

OC Actinobacillus.
 OX NCBI_TaxID=715;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IS0158;
 RX MEDLINE=96110919; PubMed=8557378;
 RA Cruz W.T., Nedialkov V.A., Thacker B.J., Mulks M.H.;
 RT "Molecular characterization of a common 48-kilodalton outer membrane
 protein of Actinobacillus pleuropneumoniae.";
 RL Infect. Immun. 64:83-90(1996).
 DR EMBL; U24492; AAC43631.1; ..
 SQ SEQUENCE 449 AA; 48604 MW; 62CBDD17A8435418 CRC64;

alignment_scores:

Quality: 1520.00 Length: 450
 Ratio: 3.979 Gaps: 3
 Percent Similarity: 84.889 Percent Identity: 66.667

alignment_block:

US-09-303-518D-127 x Q44130 ..

Align seg 1/1 to: Q44130 from: 1 to: 449

```

1 ATGATTAAATCAAAAAGGTCTAAACCTGCCCATCGCGGCGACACCGGA 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProAl 17
51 GCAAGTCATTATGACGGGCCCGTCATTACCGAAGTCGGTTCGTTGGCG 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
17 aGlnValIleHisAsnGlyAsnThrValAsnGluValAlaMetLeuGly 34
101 AAGAATATCGCGGTATGCGCCCTGTCATCAAGTCAAGGAAGCGGATGCC 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
34 LuGluThrValGlyMetArgProSerMetLysValArgGluGlyAspVal 50
151 GTCAAAAAGGCGCAAGTGTGTGTTGAAGCAAAAAGNATCCGGCGTGT 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67
201 GTTTACCGCGCGTTCAGGCAAAATCGCGCCATCCATCCGCGGAA 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
67 lPheThrAlaProIleSerGlyThrValValThrIleAsnArgGlyGlu 84
251 AGCGGTACTTTCAGTCGGTGGTATTCGCGTTCAGGCAACACGAAATC 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
84 ysArgValLeuGlnSerValValIleLysValGluGlyAspGluGlnIle 100
301 GAGTTCGAACGCTACGCGCCCGGAGCGTTGGCAAACTTAAGCGCGGANGA 350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
101 ThrPheThrArgTyrGluAlaAlaGlnLeuAlaSerLeuSerAlaGluGl 117
351 ANTNGNGCAATCTGATCAATCCGGTTGTGGACTGCGTGGTANCC 400
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
117 nValLysGlnAsnLeuIleGluSerGlyLeuThrPheAlaPheArgThrA 134
401 GTCCGTTACGAAAATCCCTCGCGTCGATCCGAGCGCTTCGCCAATCTTC 450
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
134 rgProPheSerLysValProAlaLeuAspAlaIleProSerSerIlePhe 150
451 GTCAATCGGATGACACCAATCCGTCGCGGACACCTGTGTTGTGAT 500
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProGluValValle 167
501 CAAAGAACCGCGGANGATTTCAGAGGANGTNTGCTGTTATGAGCGGTT 550
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
167 uLysGluTyrGluThrAspPheLysAspGlyLeuThrValLeuThrArgL 184
551 TG.....ACCGAGCGTAAATCCATGTGTGAAGCAGCTGCGCGCAGAC 594
|| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 euPheAsnGlyGlnLysProValTyrLeuCysLysAspAlaAspSerAsn 200
595 GTCCCG...TCTGAAATGCTGCCAATCATCGAACAACATGAAATTCGCGG 641

```

```

201 iLeProLeuSerProAlaIleGluGlyIleThrIleLysSerPheSerGl 217
||||| ||| :||| :||| :||| :||| :||| :|||
642 CCGCATCCGCGCGGTTTGTAGTGGCAGCGACATTCATTTTCATTCAGCCGG 691
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
217 yValHisProAlaGlyLeuValGlyThrHisIleHisPheValAspProv 234
692 TCGGTGCAAAACAAACCGTTTGGACCATCAATATATCAAGATGTAATTGCC 741
||||| :||| :||| :||| :||| :||| :||| :|||
234 alGlyAlaThrLysGlnValTrpHisLeuAsnTyrGlnAspValIleAla 250
742 ATCGGAGCTTTGTTTCAACAGCGCGTCTCAACACCGAGCGCGTATGCC 791
||||| :||| :||| :||| :||| :||| :||| :|||
251 iLeGlyLysLeuPheThrThrGlyGluLeuPheThrAspArgIleIleSe 267
792 TTTGGGTGTTCTCAAGTCAACAAACACCGCCTCTTTCGCTACCGCTTTGG 841
||||| :||| :||| :||| :||| :||| :||| :|||
267 rLeuAlaGlyProGlnValLysAsnProArgLeuValArgThrArgLeuG 284
842 GTGCAAAAGTATCGCAAAATTACTGCGGGCGAATGTTGACGACAGACAAC 891
||||| :||| :||| :||| :||| :||| :||| :|||
284 lYAlaAsnLeuSerGlnLeuThrAlaAsnGluLeuAsnAlaGlyGluAsn 300
892 CCGGTGATTTCCGTTTCGTTATTCAACGCGCGGATTTACACAAGCGCGCA 941
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
301 ArgValIleSerGlySerValLeuSerGlyAlaThrAlaAlaGlyProVa 317
942 CGATTATTTGGGCGCTTACCACAAATCAGATTTCCGTTATTCGAAAGGCC 991
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
317 lAspTyrLeuGlyArgTyrAlaLeuGlnValSerValLeuAlaGluGlyA 334
992 GCACAAAGAGCTGTTCCGCTGGTTCGCGCGCGAGCGGACCAATACTCC 1041
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
334 rgGluLysGluLeuPheGlyTrpIleMetProGlySerAspLysPheSer 350
1042 ATCAGCGTAGACCTCGCGCATTTCTCTGAAAAACAACTCTCAAGTT 1091
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
351 lIleThrArgThrValLeuGlyHisPheGlyLys...LysLeuPheAsnPh 366
1092 CACGACAGCGCTCAACGGTGGCGACCGCGCATGGTGGCGGATGGTACTT 1141
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
366 eThrThrAlaValHisGlyGlyGluArgAlaMetValProIleGlyAla 383
1142 ACGAGCGGTAATCCGCTAGACATCTCTGCTACCTGCTGCTTTTCGCGGAT 1191
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
383 yrGluArgValMetProLeuAspIleIleProThrLeuLeuLeuArgASP 399
1192 TTAATCGTCGCGGATACCGACGCGCGCAAGCATTTGGTTCGTTGGAATT 1241
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
400 LeuAlaAlaGlyAspThrAspSerAlaGlnAsnLeuGlyCysLeuGluLe 416
1242 GGACGAAGAAGACCTCGCTTGTGTCAGCTTCGCTGCTGCCGCGGCAATACG 1291
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
416 uAspGluGluAspLeuAlaLeuCysThrTyrValCysProGlyLysAsnA 433
1292 AATANGCGCGCTCTGCTGTAAGTGTGGAAACNTTGTGAGAGGAAGGC 1341
|| :||| :||| :||| :||| :||| :||| :|||
433 snTyrGlyProMetLeuArgAlaAlaLeuGluLysIleGluLysGluGly 449

```

seq_name: sp_bacteria:Q9X650

seq_documentation_block:

ID Q9X650 PRELIMINARY; PRT; 451 AA.

AC Q9X650;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE 50 KDA ANTIGEN PGI.

OS Porphyromonas gingivalis (Bacteroides gingivalis).

OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;

OX Porphyromonas.

NCBI_TaxID=837;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=W50;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothe L., Margetts M.,
RA Hocking D., Webb E.;
RT "Porphyromonas gingivalis polypeptides and nucleic acids."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF14076; AAD33930.1;
SQ SEQUENCE 451 AA; 49825 MW; 3E4CC9B66FE5AD74 CRC64;

alignment_scores:

Quality: 636.00 Length: 452
Ratio: 2.208 Gaps: 7
Percent Similarity: 63.717 Percent Identity: 34.071

alignment_block:

US-09-303-518D-127 x Q9X6S0

Align seg 1/1 to: Q9X6S0 from: 1 to: 451

1 ATGATTAATAAATCAAAAAGGCTAAACCTGCCATCGCCGCGGACACGC... 48
:::||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4 VallileysThrLysLysGlyLeuAlaLeuAlaLeuLysGlyLysProLe 20
49 .GAGCAAGTCAATTTATGACGGCCCGCTACCTACCGAAGTGGCTGCTG 97
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
20 uProGluMetLeuAlaGluProAlaGlnSerProThrTyrAlaValValp 37
98 CGAAGATATGCGCGTATGCGCCCTGATGAAAGTCAAGGAGCGGAT 147
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
37 roAspPheGluGlyValLeuValLysValThrAlaArgProGlyAsp 53
148 GCGGTCAAAAAGGCCAAGTCTGTTGAGACAAAGAAAGNATCGGCGCT 197
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
54 LysValArgAlaGlySerAlaLeuMetHisLysAlaTyrProGluMe 70
198 GGTGTTACCGCCGCTTTCAGCAAAATCGCCCATCCATCGCGCGC 247
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
70 tLysPheThrSerProValSerGlyGluValAlaValAlaValArgGlyA 87
248 AAAAGCGGTACTTCAGTGGTCTGATGCGGTTGAAGCAAGCAGGAA 297
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
87 laLysArgLysValLeuSerIleGluValLysProAspGlyLeuAsnGlu 103
298 ATCGAG...TTCGAACGCTACGCGCCGCGAAGCGTTGCCAACTTAAGCG 344
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
104 TyrGluSerPheProValGlyAspProSerAla.....LeuSerAl 117
345 CGANGAANTNNGNCATCTGATCAATCCGTTTGTGGACTGCGGTGC 394
: :|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
117 aGluGlnIleLysGluLeuLeuLeuSerSerGlyMetTyrPheIleL 134
395 GTANCCGTCGTCAGCAAAATCCCTCCGCTGATGCGGAGCGGTGCGC 444
::: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
134 ysGlnArgProTyrAspIleValAlaThrProAspIleAlaProArgAsp 150
445 ATCTTCGTCATGCGATGGACACCAATCCGCTNGCGCAGACCCCTGTGGT 494
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
151 IleTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspPh 167
495 TGTGATCAAGAGCGCCGANGATTTACGACGANGTNGCTGGTATTGA 544
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
167 eIleValArgGlyGluGluArgAlaLeuGlnThrAlaIleAspAlaLeuA 184
545 GCGCTTTGACGCGGTAAATCCATGTGTGTAAGGACGCTGCGCGAGAC 594
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
184 laLysLeuThrThrGlyLysValTyrValGlyLeuLysProGlySerSer 200
595 GTGCGGTCTGAAATGCTGCAACATCGCAACACATGAATTCGCGCGCCC 644
::: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
201 LeuGlyLeuHisAsnAlaGluIleValGluValHis.....GlyPr 214
645 GCATCGCGCGGTTGAGTGGCAGCACATTCATTTTCATTTGAGCGCGTGC 694

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 oHisProAlaGlyAsnValGlyValLeuIleAsnHisThrLysProIleA 231
695 GTCAACAAACAAACCGTTTGGACCATCAATATCAAGATGTAATGCGATC 744
: :|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
231 snArgGlyGluThrValTyrThrLeuLysAlaThrAspLeuIleValle 247
745 GGACGTTTGTTCACAGCGCTCTGAACACCGACGCGGTGATTCGTTT 794
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
248 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAla 264
795 GGGTGTTCTCAAGTCAACAAACCGCTCTTCGCTACCGTTTGGGTG 844
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
264 tThrGlySerAspAlaAlaHisGlyTyrValArgIleMetProGlyC 281
845 CGAAGTATCGCAATTTACTGCGGCGCAATTTGTT.....GACGCGAC 888
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
281 ysAsnValPheAlaSerPheProGlyArgLeuThrIleLysGluSerHis 297
889 AACCGGTGATTTCCGTTTGGTATTAACGCGCGGATTCACACAGGCGC 938
:|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
298 GluArgValIleAspGlyAsnValLeuThrGlyLysLysLeuCysGlu 314
939 GCACGATTTTGGGACGCTACCAATCAGATTTCCGTTTATCGAAGAG 988
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
314 sGluProPheLeuSerAlaArgCysAsnIleThrValIleProGluG 331
989 GCGCGACG...AAAAGCTGTTGCGGTTGCGCGCGACGCGGACAA 1035
|| :|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
331 LyAspValAspGluLeuPheGlyTyrAlaAlaProArgLeuAspGln 347
1036 TACTCATCAGCGTACGACCTCGCCCATTTTCCTG...AAAAACAAC 1082
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
348 TyrSerMetSerArgAlaTyrPheSerTrpLeuGlnGlyLysAsnLys 364
1083 CTTCAAGTTACGACGCGTCAACGCGTGGCGACCGCGCATGTCGCGA 1132
: :|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
364 uTyrValLeuAspAlaArgIleLysGlyGlyGluArgAlaMetIleM 381
1133 TTGGTACTTACGACGCGGTAAATCGCGCTAGACATCTGCTACCTGCT 1182
: :|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
381 exAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyr 397
1183 TTGCGCGATTTAATCGTCGCGCATACCGACGCGCGGACGATTCGGT 1232
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
398 LeuLysAlaIleAlaPheAspIleAspLysMetGluAspLeuGlyTl 414
1233 CTTGGAATTTGACGAGACGCTCGCTTGTGCGAGCTTCGCTCGCCG 1282
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
414 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspTh 431
1283 GCAATACGAATANGCGCGCTGTTGCGTAAAGTGTCTGGAACACNTT 1332
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
431 erLysIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeu 447
1333 AAGGAA 1338
|||||
448 LysGlu 449

seq_name: sp_fungi:Q9C105

seq_documentation_block:

ID Q9C105 PRELIMINARY; PRT; 1236 AA.
AC Q9C105;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE GLUCAMYLASE I (ALPHA-L-4-GLUCAN GLUCOSIDASE), EXTRACELLULAR
DE STARCH-DEGRADING ENZYME, BY SIMILARITY TO S. CEREVISIAE STAL, CONTAINS
DE CHITINASE FAMILY SIGNATURE.
GN SPAPB17.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

89 GCGACTTCGGTAATGACGGGC 69
 ||||| :|||
 911 GluThrSerSerValValgly 917

seq_name: sp_fungi:Q04051

seq_documentation_block:

ID Q04051 PRELIMINARY; PRT; 1802 AA.

AC Q04051;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE HKR1P.

GN HKR1 OR YDR420W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
 OX NCBI_taxid=4932;

RN [1]

RN SEQUENCE FROM N.A.

RA Dietrich F.S.;

RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RN SEQUENCE FROM N.A.

RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,
 RA Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
 RA Hunicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
 RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
 RA Winant A., Yelton M., Botstein D., Davis R.W.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RN SEQUENCE FROM N.A.

RA Jia Y., Cherry J.M.;

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U33007; AAB64857.1; -.

DR SGD; S0002828; HKR1.

SO SEQUENCE 1802 AA; 188945 MW; 0FF726990C7CDC51 CRC64;

alignment_scores:

Quality: 133.50 Length: 445
 Ratio: 0.588 Gaps: 20
 Percent Similarity: 51.011 Percent Identity: 22.921

alignment_block:

US-09-303-518D-127/rev x Q04051 ..

Align seg 1/1 to: Q04051 from: 1 to: 1802

1337 TCCTTCTCAANGTTCCAGCACCTTACGCAACAGCGGCGCTATTCGTA 1288
 ||||| :|||
 390 SerAspThrValSerSerSerThrAsnThrValProIySerSe 406
 1287 TTTCGCCGGGAGCGAGCTGCACAAAGGAGGCTTCTTCGTCCCAATT 1238
 :||| :|||
 406 rVal...HisSerThrPheValHis...AlaThrSerSerThrTyri 421
 1237 CCAAGCACCAACCAATGCTTGGCGGCTGTCGGTATCGCGCAGCATTAATCG 1188
 :||| :|||
 421 leSerSerSerLeuTySerSerSerProSerLeuSerAlaSerValSer 437
 1187 CGAAAAGCAGGATGAGCAGGATGCTAGGGCATTACGCGCTCGTAAGT 1138
 :||| :|||
 438 HisPheGlyValAlaPropheProSerAlaTyriIleSerPheSerVa 454
 1137 ACCAATC.....GGCACCATGGCGGCTGCCACCCTTGACGGCTG 1097
 :||| :|||
 454 lProValAlaValSerSerThrTyThrSerSerProSerAlaSerValV 471
 1096 TCGTGAACCTGAAGATTGTTTTTCAGGAATAGCGCGGGTGGTACGC 1047
 |||||
 471 alVal.....SerSerThrThrThrSerSer 472

1046 GTGATGGAGTATTTGTCCGGCTGCGCGCAACCCAGCCGCAACAGCTCTTT 997
 :||| :|||
 473ProSerAlaTyriAlaSerSerProSerValProVa 484
 996 GCTGCGGCGCTTCTTCGATAACGGAATCTGATTGGTGGTACGCTCCCAAT 947
 :||| :|||
 484 lAlaValSerSerThrTyThrSer..... 493
 946 AATCGTGGCGGCTTCTGTGAATC....GCGCCCTTCAATACCCAGCGGA 900
 :||| :|||
 494 ..ProSerAlaProAlaIleSerSerThrTyThrSerSerProSer 509
 899 ATC.....ACGGCGTGTCTGCTCAACCAATTCGCCGCGC 865
 :||| :|||
 510 AlaProValAlaValSerSerThrTyThrSerSerProSerAlaProAl 526
 864 AGTAATTCGGATACCTTC.....GCACCCAAAACGGTAC 830
 :||| :|||
 526 aAlaIleSerSerThrTyThrSerSerProSerAlaProValAlaVal. 542
 829 GCAAGAGGCGTGTGTGTTGACTTGAGAACA.....CCCAAGCAATC 786
 :||| :|||
 543SerSerThrTyThrSerSerProSerAlaProAlaIle 556
 785 AGCGCTCGGTGTTTCAGACGGCTGTTCACAAACAAACGTCGATGCAAT 736
 :||| :|||
 557 SerSerThrTyThrSerSerProSerAla.....ProValAlaVa 570
 735 TACATCTGTAATATGATGGTCCAAACGGTTTTGTTGACCGACCGGCT 686
 :||| :|||
 570 lSerSer.....ThrTyThrSerSerProSerAlaP 581
 685 CAAATGAATGAATGCTGCTGCCACTCAACCGCGGATGCGGGCGCG 636
 :||| :|||
 581 ro.....ValAlaIleSerSerThrTyThrSerSerPro 592
 635 AATTCATGTGTTTCGATGTTCAGCATTTTCAGACGGCAGCTCTCGCGC 586
 :||| :|||
 593 SerValProValAlaValSerSerThrTyThrSerSerProSerAlaP 609
 585 AGCTGCCTTACACACATGGATTTACGCTCGGTCAAA.....CGGC 545
 :||| :|||
 609 oAlaAlaIleSerSerThrTyThrSerSerProSerAlaProValAlaV 626
 544 TCAATACCAGCANACNTGCTGTAATNTCGNCGGCTCTTTTGATCACA 495
 :||| :|||
 626 alSerSerThrTyThrSerSerProSerAlaProAlaAlaIleSerSer 642
 494 ACCACAGGCTCTCGCGNAGCGGATTTGGTCCATCGCATTCACGAAGAT 445
 :||| :|||
 643 ThrTyThrSerSerProSerValProValAlaValSerSerThrTyTh 659
 444 GCGAAGCGCTCGCATCGACGGCAGGATTTTGTGAACGGACGGNTAC 395
 :||| :|||
 659 rSerSerProSerAlaProAlaAla.....IleSerSerThrTyT 673
 394 GAGCGCAGTCCCAACACCGGATTTGGATCATAGTTGCNNCANNATTCTCG 345
 :||| :|||
 673 hrSerSerProSerValProValAlaValSerSerThrTyThrSerSer 689
 344 CGCTTTAAGTTTCCCAACGCTTCGGCGGTAGCGTTGGAACCTCGATTTC 295
 :||| :|||
 690 Pro.....SerAlaProAlaAlaIleSerSerThrTyTh 701
 294 GTCGTTGCTTCAACGGCAATCACACCGACTGAAGTAGCGCTTTTTCGC 245
 :||| :|||
 701 rSerSerProSerAlaProValAlaValSerSerThrTyThrSerSerP 718
 244 CCGGATGATGGCGGATTTTGCCTGAAACNGCGCGGTAAACACACCG 195
 :||| :|||
 718 roSerAlaProAlaAlaIle.....SerSerThrTyThrSerSer 731

194 CCGGATNCTTTTGTCTCAACAGCACTTGGCTTTTTCACGGCATC 145
 |||:|||||
 732 ProSerAlaProValAlaValSerThrTyr.....ThrSerSe 745
 144 GCCTTCCTTACTTTCATCNAGGGCGCATACCGCA..... 108
 |||:|||||
 745 rProSer.....AlaProAlaAlaSerSerT 755
 107TATTCTTCGCAAGCAACGGAGCTTCGGTA 78
 |||:|||||
 755 hrTyrThrSerSerProSerAlaProValAlaVal 766
 seq_name: sp_mammal:Q29071

seq_documentation_block:
 ID Q29071 PRELIMINARY; PRT; 528 AA.
 AC Q29071;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE GASTRIC MUCIN (FRAGMENT).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
 RX MEDLINE=95275264; PubMed=775593;
 RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
 RA LaMont J.T.;
 RT "Isolation and characterization of cDNA clones encoding pig gastric
 RT mucin.";
 RL Biochem. J. 308:89-96(1995).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
 RX MEDLINE=94102478; PubMed=7506218;
 RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
 RA LaMont J.T.;
 RT "Pig gastric mucin: isolation and characterization of a cDNA clone
 RT with a novel tandem repeat.";
 RL Gastroenterology 106:200-200(1994).
 DR EMBL: U10281; AAC48526.1; -;
 FT NON_TER 1 528
 FT NON_TER 528 AA; 49907 MW; 0BFOF6879203B2EA CRC64;
 SQ SEQUENCE 528 AA; 49907 MW; 0BFOF6879203B2EA CRC64;

alignment_scores:
 Quality: 127.50 Length: 438
 Ratio: 0.585 Gaps: 19
 Percent Similarity: 49.772 Percent Identity: 21.233

alignment_block:

US-09-303-518D-127/rev x Q29071 ..

Align seg 1/1 to: Q29071 from: 1 to: 528

1343 CAGCCTTCCTTCTCAANG.....GTTTCCAGCACCTTACCCACAG 1303
 |||:|||||
 149 GlnProSerSerSerSerProProSerSerThrVal..... 162
 1302 CGGGCCNTATTTCGATTGTCGCCGGCAGACGAGCTGCACAAAGCGAGGT 1253
 |||:|||||
 163SerValGlnProSerSerSer.....SerAlaProT 174
 1252 CTCTCTGTCGAATTCGACGACCCCAATGTCGCGCGTGTGGTATCG 1203
 |||:|||||
 174 hrThrSerAlaThrSerValGlnProSer.....SerSerSer 187
 1202 CCGACGATTAATCGCGCAAAAGCAGGAGTAGGAGGATGCTAGCGGCAT 1153
 |||:|||||

188 ProProSerSerSerThrValSerValGlnThrSerSerSerSerVa 204
 1152 T.....ACGGCTCTCGTAAGTACCAATCGCACCATCGCGCGGT 1115
 |||:|||||
 204 lProThrThrSerThrThrSerValGlnPro.....SerSerSerS 219
 1114 CGCCACCGTTGACGGCTGCTGCTGAAGCTTTGAAGCTTTGTTTTCAGAAA 1065
 |||:|||||
 219 erValProThrThrSerAlaThrSerValArgSerSerSerSerS 235
 1064 TGCCCGAGGGTCGTACGCGTGATGGAGTATTTCGCGGCTGCGCGCAAC 1015
 |||:|||||
 236 ThrPro.....IleProSerThrThrSerVa 244
 1014 CCAGCCGACAGCTCTTTGCTGCGGCT...TCTTGGATAACGAAATCT 968
 |||:|||||
 244 lGlnProSerSerSerSerAlaProThrThrSerAlaThrSerVal. 260
 967 GATTGTGGTAGCGTCCCAATAATATGCGCGCTTGTGTATTCGCGCGG 918
 |||:|||||
 261GlnProSerSerSerSerThrProileProSerThr 273
 917 TTCATATACGGAACCGGAATCATCGCGGTGCTGCTGCTCAACCAATTCGCC 868
 |||:|||||
 274 ThrSerValGlnPro.....SerSerSerSerSerAlaPr 285
 867 CGCAGTAATTTGCGTACTTTCGCACCCAAACGGTACGCAAGAGCGGTG 818
 |||:|||||
 285 oThrThrSerAlaThrSerValGlnProSerSer..... 296
 817 GTTTGTTGACTTGCAGAACCCCAACCAATCATCGCGCTCGGTGTTCAGA 768
 |||:|||||
 297SerSerSerProProleSerSerThrIleSerVal..... 308
 767 CGGCTGTTGTCAAACAAACGTCGATGGCAATTACATCTTGATAATTGAT 718
 |||:|||||
 309GlnProSerSerSerSer..... 315
 717 GTTCAAACGGTTTGTGTCACCGCGCTCAATGAAATGAATGTGCG 668
 |||:|||||
 316SerProThrThrSerThr.....ThrSerV 324
 667 TGCACTCAACCGCGCGGATGCGGCGCCCGGAATCATGTGTTTCGATG 618
 |||:|||||
 324 alGlnProSerSerSerGlySerAlaProThrThrSerAlaThrSerVal 340
 617 TTGGCAGCATTTTTCAGACGCGCGTCTGCGCGAGTGCCTTACACACAG 568
 |||:|||||
 341 GlnProSerSerSerSerProProleSerSerThrIleSerValG 357
 567 GATTTTACGTCGTCACAAACGGCTCAATACCAACACNACNTCTCTGAAT 518
 |||:|||||
 357 nProSerSerSerSerSerProThrThrSerThrThrSerValGlnP 374
 517 CNTCGMCGGCTTCTTGATCACAACACACAGG..... 486
 |||:|||||
 374 roSerSerSerGlySerAlaProThrThrSerAlaThrSerValGlnPro 390
 485 TCTGCGCGNAGCGGATGCTGTCATCGCATTCAGCAAGATCGCGNACGG 436
 |||:|||||
 391 SerSerSerSerSerValProThrThrSerAlaThrSerValArgSerSe 407
 435 CTCGGCATCGACGCGGAGGATTGTCGTAACGCGGNTACGACGCGCAG 386
 |||:|||||
 407 rSerSerSer..... 410
 385 TCCACAACCGGATTCGATTCAGATTGCNNNNANTTCNTCGCGCTTAAG 336
 |||:|||||
 411SerThrProilePro 415
 335 TTTCGCAACGCTTCGGCGCGTAGGTTTCGAACTCGATTTCGTCGTTGCC 286
 |||:|||||
 416 ThrThrThrSerSerValGlnProSerSerSerSerValProThrThrSe 432


```
468 GGTGTCATCGATGACGAAGATGCGAAGCGCTCGGCATCGACGGCAG 419
:   :   :   :   :   :   :   :   :   :   :   :   :   :
918 eSerThrThrSerGluGluMetThrSerGlnGlySerThrGlnThrPro. 934
:   :   :   :   :   :   :   :   :   :   :   :   :   :
418 GGAATTTGCTGACGACGACGAGTACGAGCGGATGCTCCAAACCGGATGG 369
:   :   :   :   :   :   :   :   :   :   :   :   :   :
935 .....GlySerThrGlySerThrValThrGlnProSerThr 946
:   :   :   :   :   :   :   :   :   :   :   :   :   :
368 ATCAGATTCNNCANNANTCN...TCGCGCGCTTAAGTTTGCCACGCGTTC 322
:   :   :   :   :   :   :   :   :   :   :   :   :   :
947 ValSerAspSerThrSerGlySerThrValThrValGlySerThrGln 963
:   :   :   :   :   :   :   :   :   :   :   :   :   :
321 GGGCGCGCTAGCTGCTGCAATTCGATTCGCTGCTTCAACGGCAATCA 272
:   :   :   :   :   :   :   :   :   :   :   :   :   :
963 uGly.....SerSerSerProIleProSerThrSerGlnA 975
:   :   :   :   :   :   :   :   :   :   :   :   :   :
271 CGACCGACTGAGTACGCGCTTTTCGCGCGATGGATGGCGCGGATTTTG 222
:   :   :   :   :   :   :   :   :   :   :   :   :   :
975 snThrAsnProSerThrSer...SerGlySerSerMetSerThrGlnThr 990
:   :   :   :   :   :   :   :   :   :   :   :   :   :
221 CCTGAAACNGGC.....GGGTAAACACACCGCCCGGATN 187
:   :   :   :   :   :   :   :   :   :   :   :   :   :
991 ProGlnSerSerGlnSerThrSerProValGluSerSerThrSerGlyAl 1007
:   :   :   :   :   :   :   :   :   :   :   :   :   :
186 CTTTTCCTCTCAACAGCAGCTTGGCTTTTGTGACGGCA.....T 146
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1007 aThrSerSerSerGlySerProGlyThrThrLeuThrSerIleSerPro 1024
:   :   :   :   :   :   :   :   :   :   :   :   :   :
145 CGCCTTCCTTGACTTCATCNCAGGGCGGATACCGGCATATCTTCGCGCA 96
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1024 erProSerProSerThrIleGlySerSerGlnGlySerThrSerPro 1040
:   :   :   :   :   :   :   :   :   :   :   :   :   :
95 AGCAACGCGACT..... 84
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1041 ValValSerThrIleSerGlnGlySerThrGluThrProGlySerThrGln 1057
:   :   :   :   :   :   :   :   :   :   :   :   :   :
83 .TCGGTAATGACGGGCCGCTCAATAATGACTTCTCCGCTGTCGCCGCGCA 35
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1057 ySerThrValThrLysProSer.....ThrValSerGlySerAlaSerS 1072
:   :   :   :   :   :   :   :   :   :   :   :   :   :
34 TGGGC 30
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1072 erGly 1073
:   :   :   :   :   :   :   :   :   :   :   :   :   :
seq_name: sp_human:Q99552
seq_documentation_block:
ID Q99552 PRELIMINARY; PRT; 3570 AA.
AC Q99552;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MUCIN MUC5B (FRAGMENT).
GN MUC5B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=97166151; PubMed=9013550;
RA Desseyn J.L., Guyonnet-Duperat V., Porchet N., Aubert J.P., Laine A.;
RT "Human mucin gene MUC5B, the 10.7 kb large central exon encodes
RT various alternate subdomains resulting in a super-repeat. Structural
RT evidence for a 11p15.5 gene family.";
RL J. Biol. Chem. 272:3168-3178(1997).
DR EMBL; Z72496; CAA96577.1; -.
FT NON_TER 1
FT NON_TER 3570
SQ SEQUENCE 3570 AA; 361019 MW; DE04E4D72759312 CRC64;
```

```
alignment_scores:
  Quality: 122.50      Length: 482
  Ratio: 0.544        Gaps: 21
  Percent Similarity: 46.680      Percent Identity: 22.822

alignment_block:
US-09-303-518D-127/rev x Q99552 ..
Align seg 1/1 to: Q99552 from: 1 to: 3570

1337 TCCTTCTCAANGTTTCCAGCACCTTACGCAACAGCGGGCCNTATTCGTA 1288
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2882 SerThrAlaThrProSerSerThrProGlyThrAlaProProProLysVa 2898
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1287 TTTG.....CCGCGGACAGACGAAGCTG.....CACAAAGCGAGT 1253
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2898 lLeuThrSerProAlaThrThrProThrAlaThrSerSerLysAlaThrS 2915
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1252 CTCTTCGTCTCC.....AATTCACAG 1233
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2915 erSerSerSerProArgThrAlaThrThrLeuProValLeuThrSerThr 2931
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1232 CAACCCAAATGCTCGCGCTGCTCGATATCGCGACGATTAAATCGCGCAA 1183
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2932 AlaThrLysSerThrAlaThrSerValThrProIleProSerSerThrLe 2948
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1182 AAGCAGGCTAGGACGAGTGTCTAGCGGCATTACGCGCTCGTAAGTACCAA 1133
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2948 uGlyThrThrGlyThrLeuProGluGlnThrThr.....ThrProV 2962
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1132 TCGGCACATGGCGCGCTCGCACCGCTGAGCGCTGCTGTAACCTGAAG 1083
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2962 alAlaThrMetSerThrIleHisProSerSerThr..... 2973
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1082 AGTTTGTTCAGAAATGCGCGAGGTCTGACGCGTGATGGAGTATTT 1033
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2974 .....ProGluThrThrHisThrSerThrValLe 2983
:   :   :   :   :   :   :   :   :   :   :   :   :   :
1032 GTCCGCGCTGGCGGCAACCCAGCGGCAAGCTCTTTCTCGCGCCTTCCT 983
:   :   :   :   :   :   :   :   :   :   :   :   :   :
2983 uThrThrLysAlaThrThrArgAlaThrSerSerSerThrSerThrProSerS 3000
:   :   :   :   :   :   :   :   :   :   :   :   :   :
982 CG.....ATAACGGAAATCTGATTTGCTGCTAGCGT 954
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3000 erThrProGlyThrThrTripleLeuThrGluLeu..... 3011
:   :   :   :   :   :   :   :   :   :   :   :   :   :
953 CCCAAATAATCGTCGCGCCTTGTGTAATCGCGCGCTTCAATACCGAAC 904
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3012 .....ThrThrAlaAlaThrThrThrAlaGlyThrGlyPr 3023
:   :   :   :   :   :   :   :   :   :   :   :   :   :
903 GGAAATCACGCGGTGTCT..... 885
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3023 oThrAlaThrProSerSerThrProGlyThrThrTripleLeuThrGluL 3040
:   :   :   :   :   :   :   :   :   :   :   :   :   :
884 .....GCCTCAACCAATTCGCCGCGAGTAATT 858
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3040 euThrThrThrAlaThrThrThrAlaSerThrGlySerThrAlaThrLeu 3056
:   :   :   :   :   :   :   :   :   :   :   :   :   :
857 TCGCATACTTTTCGCACCCCAACCGGTACGCAAGAGCGCTGTTTGTGAC 808
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3057 SerSerThr.....ProGlyThrThr.....TrpIleLeuTh 3067
:   :   :   :   :   :   :   :   :   :   :   :   :   :
807 TTGAGAACCCCAACGAATACAGCGCTGGGTTCAGACGGCCTTTG 758
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3067 r...GluProSerThrThrAlaThrValThrAlaProProGlySerThrA 3083
:   :   :   :   :   :   :   :   :   :   :   :   :   :
757 CAAACAAACGCTCCGATGCAATTTACATCTTGATAATTTGATGGTCCAAACG 708
:   :   :   :   :   :   :   :   :   :   :   :   :   :
3083 laThrAlaSerSerThrGlnAlaThrAlaGlyThrProHisValSerThr 3099
:   :   :   :   :   :   :   :   :   :   :   :   :   :
707 GTTTTGTTCACCGACCGCTCAATGAATGAATGCTGCGTGCACCTCAA 658
:   :   :   :   :   :   :   :   :   :   :   :   :   :
```



```
285 rAspLeuHisIleSerSerThrProAlaAlaThrThrLeuPro..... 299
736 TTACATCTTGATAAATGATGTCCTCAACAGGTTTGTTCACCGACCGGC 687
300 ::::::::::::::::::::: 311
686 TCAATGAATGAATGTCGTGCCACTCAA...CCGGCGGATGCGGGCC 640
312 ::::::::::::::::::::: 322
639 GCGCAAT.....TCATGTGTTTCGATGTTGGCAGCATTTTCAG 602
322 nProSerHisGlyThrLeuGlyLeuSerGlyThrLeuGlyAlaTyrT 339
601 ACGGACGCTGCGGCGAGTCTACACACATGGATTTCAGCTCGGTC 552
339 hrSerThrSerValProIleSerLeuSerAlaCys..... 350
551 AAACGGCTCAATACACAGCANCTGCTGAAATCNCNCGGCTTCTTT 502
351 ::::::::::::::::::::: 365
501 GATCAACACAGGCTGCGCCNAGCGGATGGTGTCATCGCATGGA 452
365 u..... 365
451 CGAAGATGCGGACGCTCGCATCGACGCGAGGATTTTGTGAACGGA 402
366 ::::::::::::::::::::: 378
401 CGGNTACGACGACGCTCCAAACCGGATTTGGATCAGATTGCNCGNNAN 352
379 HisGlySerSerThrProIleAlaProValPheThrAlaLeuProSerPh 395
351 TTCTNCGCCCTTAAGTTTCCCAACGCTTCGGCGGCTAGCGTTCAACT 302
395 eThrSerLeuThrAsnAsnPheProLeuThrGlyAsnProSerLeuAsnP 412
301 CGATTCGCTGTTGCTTCAACGGCAATCAGCAGCGACTGAAGTAGCGGC 252
412 roSerValSerLeuProGlySerLeuIleAlaThrSerSerThrAlaAla 428
251 TTTTCG.....CCGGATGATGCGGCGGATTTTG..... 222
429 ThrSerThrSerLeuProHisProSerSerThrAlaAlaValLeuSerG1 445
221 .....CCTGAACNGCGG..... 207
445 yLeuSerAlaSerAlaProValSerAlaAlaProPheProLeuAsnLeuS 462
206 .....GTAAACACCGCGCGGATNCITTTTGTCT 177
462 erThrAlaValProSerLeuPheSerValThrGlnGlyProLeuSerSer 478
176 TCAACACGACTTGGCTTTTTCACCGCATCG.....CCTTCCTTGAC 133
479 SerAsnLeuSerTyProGlyPheSerValSerAsnThrProSerValTh 495
132 TTTTCATCAGGGCGCATCGGCATATTCTCGCCAAAGCAACGCGACTT* 83
495 iProAlaLeuProSerPheProGlyLeuGlnAlaProSerThrValAlaA 512
82 CGGTA.....ATGACGGCGCGTCA 63
512 laValThrProLeuProValAlaAlaThrAlaProSer 524
seq_name: sp_fungi:094317
seq_documentation_block:
ID 094317 PRELIMINARY; PRT; 534 AA.
AC 094317;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
```

```
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 53.4 KDA SERINE-RICH PROTEIN C215.13 IN CHROMOSOME II
DE PRECURSOR.
GN SPBC215.13.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -!- SIMILARITY: TO YEAST STAL.
DR EMBL: AL033534; CAA22127.1; -.
KW Hypothetical protein; Glycoprotein; Signal.
FT SIGNAL 1 14
FT CHAIN 15 534
FT CARBOHYD 31 31
FT CARBOHYD 426 426
FT CARBOHYD 534 AA; 53361 MW; D62DC15C80A698E4 CRC64;
SQ SEQUENCE 534 AA; 53361 MW; D62DC15C80A698E4 CRC64;

alignment_scores:
Quality: 121.00 Length: 420
Ratio: 0.582 Gaps: 15
Percent Similarity: 49.524 Percent Identity: 19.762

alignment_block:
US-09-303-518D-127/rev x 094317 ..
Align seg 1/1 to: 094317 from: 1 to: 534

1280 GGGCAGACGAAAGTGCACAAAGCGAGGTCTTCTCGTCCAATCCAAAGCA 1231
67 GlyGlyThrProThrTyrTyrGlyTyrAlaThrProThrSerSerG1 83
1230 ACCCAATCCTTTCGGCGTGTGCGGTATCGCGACGATTAATTCGCGCAAA 1181
83 uProSerIlePheSerGluSerAlaThrPro.....S 94
1180 GCAGGTAGGACGATGCTAGCGGCATTACGCGCTCGTAAGTACCAATC 1131
94 erGluThrAsnSerTyrSerSerProValSer..... 104
1130 GGCACCATGGCGCGTCCGCCACCGTTGACGGCTGTCGTGAACATTGAAGAG 1081
105 .....SerTyrSerAspProAlaThrSerGlnLeuProSerSerTh 118
1080 TTTGTTTTTCAGGAAATGCCGAGGTCGTACGCGTATGGATTTGT 1031
118 rSerPhePheSerPro.....ThrSerSerGluTyrThrP 130
1030 CCGGCTCGGCGCAACCCAGCGCAACAGCTCTTCTGCTCGCGCTTCTTCG 981
130 ro.....SerSerThrGluSerSerSerLeuLeuAspProSerSer 143
980 ATAAGCGAAATCTGATTGTGTGTAGCTCCCAATAATCTGCGCGCCTTG 931
144 ValSerSer..... 146
930 TGTATTCGCGCGCTTCATACCGAACCGGAAATCACGCGTGTCTCGCT 881
147 AlaIleLeuProSerSerThrSerValGluValSerIleSerSerS 163
880 CAACCAATTCGCGCGCAGTAATTTGCGATACCTTTCGACCCCAACGGA 831
163 erLeuSerSer..... 167
830 CGCAGAGCGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 781
```

```

168 ..... AspProLeuThrSerSerThrPh 175
780 CTGGTCTTTCAGAGGCTCTTCCAAACAAACGTCGATGGCAATACAT 731
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
175 eSerSerLeuSerSerThrSerSerGlnPro...SerValSerS 191
730 CTTGATAAATGATGCTCCAAACAGTTTGTGTTGTCACCGCGCTCAATG 681
    || : : : : : : : : : : : : : : : : : : : : : : : : :
191 ex.....ThrSerSerThrPhe..... 197
680 AATGATGTCGTGCGCACTCAACCGCGCGGATCGCGCGCGCAATTC 631
198 .....SerSerAlaAlaProThrSerTh 205
630 ATGTGTTTCGATTTGGCAGCATTTTCA.....GACGCGACGTGCGC 587
    : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
205 rSerSerSerTyrLeuSerSerSerValValSerSerSerSerP 222
586 CAGCTCCCTTACACATGATTTTACGCTCGGTCAACGCGCTCAATACC 537
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
222 roSerSerSerSerSerThrLeu.....Thr 231
536 AGCANACNTCGTGAATCNTCGCGCTTCTTGTGATCAACACACAGG 487
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
232 SerSerSerLeuSerSerSerThrProSerThrSerSerSerSe 248
486 GTCTGCGCGNACGAGTGGTGTCTCATGCGATGACGAGATGGCGAAG 437
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
248 rSerThrSerSerSerLeuSerSerSerSerSerSerThrAlaSerS 265
436 GCTCGGCATCGACGCGAGGATTTGTGACGGGACGGNTACGACGCA 387
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
265 erSerSerSerSerSerSerSerSerSerSerSerSerSerSer... 280
386 GTCCACAAACCGGATGGATGATGATGCTGCNNCNTTCNTCGCGCTTAA 337
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
281 .....SerSerProThrSe 285
336 GTTTCGCAACGCTTCGGCGGTACGCTGCAATCGATTTTCGTGCTGC 287
    : : : : : : : : : : : : : : : : : : : : : : : : :
285 rThrSerSerThrLeSerSer.....SerSerSerSerSerSerP 300
286 CTTCAACGCGCATACGACC.....GACTGAAGTACGCGC 252
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
300 toThrSerThrSerSerThrSerSerSerSerSerSerSerSer 316
251 TTTTCGCGCGATGATGCGCGGATTTGCTGAAACNGCGCGGTAA 202
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
317 PheSerSerThrLeuSerSerSerSerMetSerSerSerSerPheSe 333
201 CACACGCGCGGATNCTTTTGTCTTCAACAGCACTGGCCCTTTTGA 152
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
333 rSerSerPro.....ThrSerSerSerThrSerSerSerS 347
151 CGGATCGCTCTCTGACTTTCATCAGGCGCCATACCGGCATATCT 102
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
347 erSerSerProSerSerSerPheSerSerThrThrSerSerSerLys 363
101 TCGCCACGACGCGACTTCGGTAAATGACGGCGCGCTCATAAATGACTG 52
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
364 SerSerSerPheSerSerThrValSerSerSerSerSerThrSerSe 380
51 CTCGGGTCTG 42
380 rSerThrLeu 383

```

seq_name: sp_bacteriap:Q9WY86

seq_documentation_block:

ID Q9WY86

AC Q9WY86; PRELIMINARY;

DT 01-NOV-1999 (TReMBLrel. 12, Created)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ELECTRON TRANSPORT COMPLEX PROTEIN, PUTATIVE.
GN TM0244.

OS Thermotoga maritima.

OC Bacteria; Thermotogales; Thermotoga.

OX NCBI_TaxID=2336;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MSB8 / DSM 3109;

RX MEDLINE=99287316; PubMed=10360571;

RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,

RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,

RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,

RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,

RA Heideberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,

RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.; White O.,

RT "Evidence for lateral gene transfer between Archaea and Bacteria from

genome sequence of Thermotoga maritima.";

RL Nature 399:323-329(1999).

DR EMBL; AF001708; AAD35335.1; -.

DR TIGR; TM0244; -.

DR InterPro; IPR001450; 4Fe4S_ferredoxin.

DR InterPro; IPR001949; Complex1_51K.

DR Pfam; PF01512; Complex1_51K; 1.

DR Pfam; PF00037; fer4; 1.

DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; 2.

KW Iron-sulfur; Complete proteome.

SQ SEQUENCE 451 AA; 49515 MW; 7B2D7272FC6E418 CRC64;

alignment_scores:

Quality: 120.00 Length: 456

Ratio: 0.569 Gaps: 18

Percent Similarity: 46.272 Percent Identity: 20.175

alignment_block:

US-09-303-518D-127 x Q9WY86

Align seg 1/1 to: Q9WY86 from: 1 to: 451

49 GAGCAAGTCATTATGACGGCCGTCATTACCGAAGTCGCGTTCGTTGG 98

35 AsplysProIleGluArgAlaProLeuProGlnLysValPheValPheLe 51

99 CGAAGATATGCGCGTATCGGCCCTTCGATGAAAGTCAAGGAGCGATG 148

51 uSerAsnHisAlaGlyAsnProAlaLysProValValSerProGlyAsp 68

149 CCGTCAAAAAAGCCCAAGTCGCTTTGAAGACAAAAAGNATCCGGCGTG 198

68 luValLysThrGlyGlnValIleGlyGluProGluGlyPheIleSerAla 84

199 GTGTTTACCGCGCNGTTTCAGCAAAATCGCCGCTCCATCCATCGCGCA 248

85 TyrLeuHisSerProValThrGlyArgValLeuGluIle..... 97

249 AAAGCGCGTACTT.....CAGTCGTCGTGA 274

98 .LysLysIleLeuHisProIleLeuGlyLysProIleGluAlaIleValI 114

275 TTGCGCGTTGAAGGCAACGACGAAATC.....GAGTTC 306

114 leGluArgThrSerAspAspGluTrpValHisIleGluThrGlyAspPhe 130

307 GAACGCTAGCGCCCGCAAGCGTTGCAAACTTA.....AGCGCGCA 347

131 GluArgMetSerLysGluIleLeuGluIleIleLysLysAlaGlyI 147

348 NGAANTNNGNCAATCTGCATCAATCCGTTTGTGGACTCGGTCGCGTA 397

147 e.....ValGlyLeuGlyAlaMetPheP 156

398 NCCGTCGCTTCAGCAAAATCCCTCGCGTCGATCGCGAGCCGTCGCCATC 447

```
156 roThrHisValLysLeuSerProProGluLysLysValAspThrLeu 172
    :::      |||      :::      :::      :::::
448 TTCGTCATCGATGACACCAATCCGCTNGCGGACGCCCTGTGTGT 497
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
173 lIeValAsnGlyAlaGluCysGluProValLeuThrIleAspHisArgLe 189
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
498 GATCAAGAGCGCGGANGATTTCAGAGGANGTNTGCTGTATTGAGCC 547
    ::|||:::  |||  ::|||:::  ::  ::|||:::  ::
189 umetLeuGluArgAlaGluAspIleLeuGlnGlyIleLeuMetL 206
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
548 GTTTGACCGAGCGTAAATCCATGTGTGTAAGCA...GCTGGCGCAC 594
    ::|||:::  ::  |||  ::|||:::  ::  ::|||:::  ::
206 ysVal.....LeuGlyValGlnLysAlaValValGlyValGlu 218
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
595 GTCCGCTCGAATGCTCCCAACATCGAACAACATGATTCGGCGGCC 644
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
219 SerAsnLysMetAspAlaTyrHisAsnLeuLysLysValPheLysGly.. 234
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
645 GCATCGCGCGGTTTGAGTGGCAGCACACATTCATTTCATTGAGCGGTCG 694
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
235 .TyrProValAspValAlaLeuLeuArgThrLysTyr.....ProGlnG 249
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
695 GTCAACAACAAA...ACCGTTTGGACATC..... 720
    ::|||:::  |||  ::|||:::  ::  ::|||:::  ::
249 lYalaGluLysGlnLeuIleTyrAlaIleThrGlyArgMetValProArg 265
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
721 .....AATTATCAAGATGT 734
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
266 GlyClyLeuProMetAspValGlyValValValGlnAsnValGlyThrCy 282
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
735 AATTGCCATCGGACGTTGTTTGCACAGCGGCTCTGAACACGAGCGCG 784
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
282 sValAlaValLysGluAlaValValAspGlyLysProLeuValGluArg 299
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
785 TGATTGCTTGGTGTCTCAAGTCAACAACACCGCTCTGCGTACC 834
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
299 lYMetThrValSerGlyAspAlaValLysAsnGlnLysAsnLeuIleVal 315
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
835 GTTTTGGTGGTGAAGATGTCGAATTTACT.....GCGGGCGAATTG 878
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
316 ArgIleGlyThrProValLysAspValIleAspTyrCysGlyGlyIleAs 332
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
879 TGAGCGAGACACCGGCGTGTATTCGGTTCGGTTCGATTCACAGCG...GCCA 925
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
332 pGluAsnThrGluArgValIleLeuGlyGlyProMetMetGlyIleSerI 349
    ::|||:::  ::  ::|||:::  ::  ::|||:::  ::
926 TTACACAAGCGCGCAGCATTTATTGGGACGCTACCACAAATCAGATTCC 975
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
349 leThr.....AsnLeuAspIle 354
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
976 GTTATCGAAGAGCGCGCAGACGAGCTGTTGCGCTGGTTCGCGCGCA 1025
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
355 ProValMetLysGlyThrSer..... 361
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
1026 GCCGGACAATCTCCATCAGCGGTACGACCCCTCGCCATTCCTCAAAA 1075
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
362 .....GlyIleThrAla.....PheLeuProL 369
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
1076 ACAAACTCTCAAGTTTCACGACAGCGCTCAACGGTGGCGACCGCGCATG 1125
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
369 ysLys.....SerArgProGlnLysProCys 377
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
1126 GTGCCGATTGGTACTTACGAGCGGTAATGCCGCTAGACATCTCGCTAC 1175
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
378 lIeArgCysSerGluCysValGlnValCysProMetAsnLeuProTy 394
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
1176 CCTCTTTTGGCGATTTAATTCGTCGCGGATACCGACAGCGCGCAAGCAT 1225
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
394 rLeuLeuTyrLeuLeuSerThrLysArgLysTyrAspGluAlaValGluA 411
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
1226 TGGGTTGCTTGGAAATGGAGAGAGACCTCGCTTGTGACGCTTCGTC 1275
    ::|||::  |||  ::|||:::  ::  ::|||:::  ::
```

```
411 snGlyLeuMetAspCysIleGlu.....CysGlySerCysThrTyrThr 425
1276 TGCCCGGGCAATAACGAA 1293
    |||||:::|||||
426 CysProSerLysIleGlu 431
    |||||:::|||||
seq_name: sp_human:Q14881
seq_documentation_block:
ID Q14881 PRELIMINARY; PRT; 622 AA.
AC Q14881;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MUCIN (FRAGMENT).
GN MUC5B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TRACHEOBRONCHIAL MUCOSA;
RA Dessen J.L., Guyonnet-Duperat V., Porchet N., Aubert J.P., Laine A.;
RT "Human mucin gene MUC5B: the 10.7 kb large central exon encodes
RL various alternate subdomains resulting in a super repeat.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DDJB databases.
DR EMBL; X74955; CAA52910.1; -.
FT NON_TER 1
FT NON_TER 1
FT NON_TER 622
FT NON_TER 622
SQ SEQUENCE 622 AA; 61786 MW; 4FC85A52F50D57E5 CRC64;

alignment_scores:
Quality: 118.00 Length: 357
Ratio: 0.674 Gaps: 20
Percent Similarity: 49.020 Percent Identity: 23.810

alignment_block:
US-09-303-518d-127/rev x Q14881 ..
Align seg 1/1 to: Q14881 from: 1 to: 622

1025 TGCGCGCAACACCGAGCGAGCTCTTGTGCGGCGCTTCTCGATAAC 976
    |||||:::|||||
59 CysProSerThrProAlaThrSerSerThrAlaMetProSerSer...Th 74
    |||||:::|||||
975 GGAATCTGATTGTGTAGCTGCCAAATAATCGCGCGCTTGTGTAA 926
    |||||:::|||||
74 rProGlyThrThrTrp.....IleL 81
    |||||:::|||||
925 TGCGCGCTTCAATACGAGCAACCGGAAATCAGCGGTTGTCGCTCAAC 876
    |||||:::|||||
81 eutThrGluLeuThrThrAlaThrThrAlaSerThrGlySerThr 97
    |||||:::|||||
875 AATTCGCGCGCAGTAATTTGGGATACTTTCGACCCCAACACGGTACGCA 826
    |||||:::|||||
98 AlaThrProSerSerThrProGlyThr...AlaPro..... 108
    |||||:::|||||
825 GAGCGGTGGTTGTGACTTGAGAACACCCCAACCAATCAG..... 783
    |||||:::|||||
109 .....ProProLysValLeuThrSerProA 117
    |||||:::|||||
782 .....CGCTCGGTGTTTCAGACGCGCTTGCACAAACAAACGTCG 744
    |||||:::|||||
117 laThrThrProThrAlaThrSerSerLysAlaThrSerSerSerPro 133
    |||||:::|||||
743 ATGGCAATTACATCTGATAATATGATGTCACAAACGGTTTGTGTGACC 694
    |||||:::|||||
134 ArgThrAlaThrLeuProValLeuThrThrAla..... 146
    |||||:::|||||
693 GACCGGCTCAATGAATGAATGTGCTGCCACTCAACACCGCGCGATCG 644
    |||||:::|||||
```

```

147 .ThrLysSerThrAlaThrSerValThrProIleProSerSerThrLeuG 163
643 GCGCG.....CCGAATTCATGCTTTTCGATGTGGCAGCATTT 606
163 lyThrThrGlyThrLeuProGluGlnThrThrProValAlaThrMet 179
605 TCA.....GACGGCAGCTCGCGCAGCTGCCTTACACACATGATTT 562
180 SerThrIleHisProSerSerThrProGluThrHisThrSerThrVa 196
561 ACGTCGGTCAACCGCTCAATACCCAGCANACNTCGTCTGAAATNCGN 512
196 lLeuThrThrLysAlaThrThr.....ArgAlaThrSerSer. 209
511 CGGCTCTTTGATCACAACACAGGCTCGCCGNCAGCGATGTGTC 462
210 .....ThrSerThrProSerSerThrProGlyThrThrTrp 221
461 ATCGCATTCAGCAAGATGGCGACGCTCGGCATCGACGCGGATTTT 412
222 Ile...LeuThrGluLeuThrThrAlaAlaThrThrAlaGly..... 235
411 GCTGAACGGCAGGNTACGCGCAGCTCCACAAACCG.....GATTGGA 368
236 ....ThrGlyProThrAlaThrProSerSerThrProGlyThrThrTrp 251
367 TCAGATTGNCNNANTNCTGCGCCGCTTAAGTTGCCAACGCTTCGGGC 318
251 leLeuThrGluLeuThrThrAlaThrThrAlaSerThrGlySer 267
317 GCGTAGCTTCGAACCTCGATTGCTGCTGCTTCAACGGCA....ATCAC 271
268 .....ThrAlaThrLeuSerSerThrProGlyThrThrTrpIleLe 281
270 GACCGACTCAAGTACGCGCTTTTCG.....CGCGATCGATGGCG 230
281 uThrGluProSerThrThrAlaThrValThrProThrGlySerThrA 298
229 CGATTTCGCTGAACNGCGCGGTAAACACACCGCCGCGATNCTTTTG 180
298 laThrAlaSerSerThrGlnAlaThrAlaGlyThrPro.....HisVal 312
179 TCTTCAACAGACATGGCTTTTGTAGC.....GCATCGCCTTC 139
313 SerThrThrAlaThrThrProThrValThrSerSerLysAlaThrProSe 329
138 CTTGACTTTCATCAGGCGCATACCGGCATATTCTTCGCCAAGCAACG 89
329 r.....SerSerProGlyThrA 335
88 CGACT.....TCGGTAATGACGGGCCCTCATATAATG 57
335 laThrAlaLeuProAlaLeuArgSerThrAlaThrThrProThrAlaThr 351
56 ACTTGCTCGGCTCGCCGCG 36
352 SerPheThrAlaIleProSer 358
seq_name: sp_archaeap_Q9HNR5
seq_documentation_block:
ID Q9HNR5 PRELIMINARY; PRT; 494 AA.
AC Q9HNR5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE VNG1983H.
DE VNG1983H.
GN VNG1983H.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]

```

```

SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leitthausser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA "Genome sequence of Halobacterium species NRC-1.";
RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005093; AAG20155.1; -.
KW Complete proteome.
SQ SEQUENCE 494 AA; 54739 MW; 55593A8FD31A72F3 CRC64;

```

```

alignment_scores:
  Quality: 117.00      Length: 529
  Ratio: 0.574        Gaps: 32
  Percent Similarity: 38.563  Percent Identity: 25.520
alignment_block:
US-09-303-518D-127 x Q9HNR5
Align seg 1/1 to: Q9HNR5 from: 1 to: 494
66 CGGCGCCGTCATTACCGAAGTCGCTGCTGGCGAAGATATCCCGTA 115
20 ArgAlaArgArgArgSerArgAla..... 28
116 TCGCGCCCTNGATGAAAGTCAAGGAGCGATGCCGTCAAAAAGGCCAA 165
29 .....GluArgAspProGlyValArg.....G 36
166 GTGCTGTTGAAGACAAAAGNATCCGGCGTGTGTATTACCGCGCNGT 215
36 lyAlaValLeuGlyAlaArgThrHisAspArgGlyGlyValArgAla... 51
216 TTCAGGCAAAATCCCGCCAT.....CCATCGGCGGCAAAAGCGCTAC 259
52 ...ArgArgGlyArgArgHisArgProProLeuArgThrAlaArgG1 67
260 TTCAGTCGTCGTGATTGCCGCTTCAAGGCAA.....CGAC 294
67 yArgThrGlyValAspSerAlaAlaArgAspAlaAlaProValProA 84
295 GAAATCGAGTTCGA..... 308
84 rgThrAlaValArgAspGlyGlyGlnThrArgGluCysProSerArg 100
309 .....ACGCTAGCGCCGCGAAG 325
101 GlyProValSerValGlyValThrAlaGluProProValArgArgPr 117
326 CGTTGGCAAACTTAAGCGCGGANGA..... 350
117 oValGlyArgLeuLeuArgValLeuArgGlyAlaArgValAspA 134
350 ..... 350
134 rgAlaGlyAlaValArgGlyArgAlaArgAlaAspThrAlaSerGlyArg 150
351 .....ANTNNGNCAATCTGATCCAAATCCGTTT...GTGACTGCGC 391
151 ProProArgArgGlyGluThrArgProHisArgAlaAspValGlyCysG1 167
392 T.....GCGTANCCGTCG 405
167 yArgProValCysAspArgValGlyAlaAsnArgProAlaGluProArgH 184
406 TTCAGCAAAATCCCTCCGTCGATGCCGAGCGCTTCGCCATCTT..... 449
::: :::: ||||| ||| ||::|||

```

184 isArgLeuGlnHisCysArgArgLeuArg.....ArgArgLeuValAla 198
450CGTAATGCCAT...GGACACCAATCC... 473
199 LeuGlnAspAlaLeuGlyProArgAlaGlyAspArgGlyHisGlyAlaVa 215
474GCTNGCGCGACGCCCT 489
215 lleuProGluValArgLeuArgValProArgValArgGlyArgProv 232
490 GTGGTGTGTGATCAAAAGACGCGGANGATTTTCAGACGANGTNGCTGGT 539
232 alAlaValAspLeuArgProArgGlyArgAspThrArgProAla... 247
540 APTGAGCCG.....TTTGACGAGCGGTAAATCCATGTGT 574
248 LeuAspProGlyArgThrAlaAlaArgAspArgProGluArgPro..... 262
575 GTAAGCAGCTGGCGCAGCGCTGCTGAAATGCTGCCAATCGAA 624
263ArgArgArgPro.....ArgValGlyGlyProArg 273
625 ACACATGAATTCGGCGCGCGCCATCCGCCGCTTGTAGTGGCAGCACAT 674
273 rgProLeuValArgArgProLeuArg..... 281
675 TCATTTTCATTCACCGGTCGG.....TGCACAAACAAACCGTTT 712
282HisLeuValAspArgProArgAspCysArgGlySerArgLe 296
713 G.....GACCATCAATTA...TCAAGATGTAATGCCATCGGA 747
296 uArgValAlaAlaArgHisAlaValAlaGlyArgAspSerAlaProArg 313
748 CGTTTTCACACAGCGCGCTGACACCGCGCGGTGATTCGTTTGGG 797
313 lyAlaAlaGlyArgArgPro.....HisArgGlyArgArgGlyLeuAla 327
798 TGGTTCTCAAGTCAACAAACACCGCTCTTGGTACCGTTTGGGTGCGA 847
328 GlyAlaArgGlnArgGlyValProGlyLeuArgArgArgAlaGlyArgLe 344
848 AAGTATCGCAATTAATCTGCGGCGAATGGTTGACGACGACGCGCGT 897
344 uArgLeuAlaAlaArgArgAlaMetGly.....ArgProGlyArgG 359
898 ATTTCCGGTTCGTATTGACGCGCGGTATACACAGGCGCGCACGATTA 947
359 lyHisGlnPheGlyLeuGlyArgArg.....ArgProAlaArgLeu 372
948 TTT.....GGGAGGCTACCAATCAGATTTCCGT..... 977
373 AspAspProGlyProArgProArgAlaProValArgAspValSerArg 389
978TATCGAAGAGCGCGGACGACGACGAGTGTTCGCG 1011
389 gArgArgThrArgGlyArgArgArgProProGlnArgAlaAlaArgArg 406
1012 TGGTTTCGCGCGCGCGGACAAATACTCCATCAGCGTACGACCCCTCGG 1061
406 roGlyGlyAlaArgGlnCysArg..... 413
1062 CCATTTTCCTGAAACAACTCTTCAAGTTTCAC...GACAGCGGTCAACG 1108
414ProSerLysSerAlaLeuProLeuAlaValThrGlyArgProAr 428
1109 GTGGCGACCGCGCGATGGTGGCGGATTTGACTTACAGCGCGT...AATG 1155
428 gGlyArgSerLeuAlaGlyThr.....ArgAlaArgThrArg 441
1156 CCCTAGACATCTCCTACCTACCTGCTTTTGGCGGATTTAATGCTGCGGCA 1205
441 lyAlaSerArgAlaAlaArgProAla.....SerSerArgArgArg 454

1206 TACCGACAGCGCGCAAGCATTTGGTTGCTTGAATTCGACGAGAGACC 1255
455 SerArgProAlaAspThrSerGly.....SerPr 464
1256 TCCTTTTGTGTCAGCTTCGTCGCCGCGCAATACGA 1292
464 OAArgThrAlaArgAlaGluAlaValGlyValValArg 476
seq_name: sp_invertebrate:Q9N5K0
seq_documentation_block:
ID Q9N5K0 PRELIMINARY; PRT: 1203 AA.
AC Q9N5K0;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 120.1 KDA PROTEIN.
GN H43E16.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Miller N., Carter T.;
RT "The sequence of C. elegans cosmid H43E16.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006669; AAF39909.1;
DR InterPro: IPR000082; SEA.
KW SMART; SM00200; SEA; 1.
KW Hypothetical protein.
SQ SEQUENCE 1203 AA; 120104 MW; B04CADC2BA74713B CRC64;
alignment_scores:
Quality: 116.50 Length: 497
Ratio: 0.507 Gaps: 19
Percent Similarity: 46.278 Percent Identity: 21.529
alignment_block:
US-09-303-518D-127/rev x Q9N5K0 ..
Align seg 1/1 to: Q9N5K0 from: 1 to: 1203
1343 CAGCCTTCCTTCCTCAANGTTTCCAGCACCTTACGCAACAGCGGCCNTA 1294
411 GlnGlySerProSerThrValSerGlnThr.....ThrProGlyValSe 425
1293 TTCGTATTTCGCCGCGCAGACGAGAGCTGCACAAAGCGAGGTCTTCGT 1244
425 rSerAlaSerThrGlyMetThrThrSerGlnAlaSerLeuArgSerThrG 442
1243 CCAATTC.....AGCAACCAATCTTCGCGCTGTCGTATCGCG 1200
442 InAsnSerValThrSerThrProGlySerPheThrValThrSerThrPro 458
1199 ACGATTAATCGCGCAAAAGCAGGCTAGGAGGATGTCTAGCGCATAC 1150


```
764 oGluProAlaGlu.....LysSerArgSerArgArgArgArgSera 779
112 GGTATGCGCCCTNGATGAAGTCAAGGAAGCGCATCGCTCAAAAAGG 161
779 laSerSerProArgThrLysThrThrSerArgArgGlyArgSerProSer 795
162 CCAAGTCTGTTTGAAGACAAAAAGNATCGGCGCTGTTGTTACCGCGC 211
796 ProLysPro.....ArgGlyLeuGlnArgSe 804
212 CNGTTCAGCAAAATCGCGCCCATCCATCGCGCGGAAAGCGCTACTT 261
804 r..... 804
262 CAGTCGTCGTGATTGCGTGAAGCAACGAGCAAAATCGAGTTCGAACG 311
805 .....ArgSerArgSerArgGlnLysThrArgThr 815
312 CTACGCGCCGAAGCTTGGCAAACTTAAGCGCGCGGANGAANTNNGNCA 361
816 ThrArgArgArgAsp.....ArgSerGlySerSerGln 826
362 ATCTGATCAAAATCGCGTGTGACTGCGCTCCGTANCCGTCGCTCAGC 411
826 nSerThrSerArgArg.....ArgG 833
412 AAAATCCCTCGCGTCGATCGCGAGCGGTCGCCATCTTCGTAATCGCAT 461
833 InArgSerArgSerArgSerArgValThrArgArgArgGlyGlySer 849
462 GGACACCAATCCGCTNGC..... 479
850 GlyTyrHisSerArgSerProAlaArgGlnGluSerSerArgThrSerSe 866
480 .....GCCAGA.....CCCTGTGTTGTGATCAAAAGA 507
866 rArgArgArgArgGlyArgSerArgThrProProThrSerArgLysArgS 883
508 GCGGCGGANGATTTCAGACGANGNTGCTGTTATGAGCGGTTGACCGA 557
883 erArgSer.....ArgThrSerProAla.....ProTyrLysArg 894
558 GCGTAAATCCATGTGTGAAGCAGCTGCGCAGACGTCGCTCTGAAA 607
895 Ser.....ArgSerArgAla.....Se 900
608 ATGTCGCAACATGCAACACATCAATTCGCGCGCGCGCATCGCGCGT 657
900 rProAlaThrHisArgArgSerArgSerArgThrProLeuLysSerArg. 916
658 TTGAGTGGCAGCGCACATTCATTTCATTGAGCGCGTGGCGCAACAAAC 707
917 .....ArgArgSerArgSer 921
708 CGTTTGGACCATCAATATCAAGATTAATTCGATCGGACGTTGTTG 757
922 ArgThrSerProValSerArgArgArgSerArgSerArgSerValTh 938
758 CACAGCGCGCTGTAACACGCGCGCGCTGATTCGTTGGTGTCTCAAA 807
938 rArgArgArgSerArgSerArgAlaSerProValSerArgArgArgSer 955
808 GTCAACAAACCGCTCTTGGTACCGTTTGGTGGGAAGTATCGCA 857
955 rSerArgThrProProValThr..... 962
858 AATTACTGCGGCGAATGTTGACGCGACAGCAACCGCGTATTCGCGTT 907
963 .....ArgArgArgSerArgSerArgThrPr 971
908 CGGTATTGAACGCGCGAT.....TACACAAGCGCG 939
971 oThrThrArgArgArgSerArgSerArgThrProProValThrArgArg 988
```

```
940 CACGATTATTGGGACGCTACCAATACAGATTTCGTTATCGAAGG 989
988 rgSerArgSerArgThrProProValThr.....ArgArgArg 1000
990 CCGCAGCAAAAGAGCTGTCGGCTGGG.TTCGCGCGCAGCCGCAAAATAC 1038
1001 SerArgSerArgThrSerProLysThrArgArgSerArgSerArgTh 1017
1039 TCCA.....TCAGCGCTACGACCCCTCGGCCATTTCCT 1070
1017 rSerProValThrArgArgArgSerArgThrSerProValThr 1034
1071 GAAAAACAACCTTCACGTTACAGAGCCCTCAACGCTGCGCGACCGC 1120
1034 rgArgSerArgSerArgThrSerProValThr.ArgArgArgSerAr 1050
1121 CCATGTCGCGGATTGCTACTACGAGCGGTAATGCCCTAGACATCTG 1170
1050 gSerArgThrProProAlaLysArgArgSerArgSerArgThrProL 1067
1171 CCTACCGCTCTTTCGCGGATTTAATCGTCGCGGATACCGACGCGCA 1220
1067 euLeuPro.....ArgLysArgSerArgSerArgSer 1077
1221 AGCATTTGGTTCGTTGGAATTGGAGCAAGAACCTCGC 1259
1078 ProLeuAlaLys.....ArgArgArgSerArg 1086
seq_name: sp_human:060382
seq_documentation_block:
ID 060382 PRELIMINARY; PRT; 1791 AA.
AC 060382;
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE KIAA0324 (FRAGMENT).
GN KIAA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,
RA Teng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RT "Sequencing of Human Chromosome 16p13.3";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ricke D.O.;
RT "Large Scale Sequence Analysis and Annotation with the Sequence
RT Comparison Analysis (SCAN) System.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004493; AAC08453.1; -.
FT NON_TER 1
SQ SEQUENCE 1791 AA; 191306 MW; 3A7B5530AE95F3E CRC64;
```

alignment_scores:

Quality:	116.50	Length:	447
Ratio:	0.613	Gaps:	19
Percent Similarity:	42.506	Percent Identity:	24.161

alignment_block:

US-09-303-518d-127 x 060382 ..

Align seg 1/1 to: 060382 from: 1 to: 1791

12 CAAAAAAGCTCTAAACCTGCGGCGGACCGGCAAGTCATTT 61

964ArgArgArgSerArgSerArgThrPr 972
908 CGGTATTGACGGCCGAT.....TACACAGCGCG 939
972 othrThrArgArgSerArgThrProProValThrArgArg 989
940 CAGGATTATTGGGACCGTACCACAAATCAGATTTCCGTTATCGAAGAGG 989
989 rgSerArgSerArgThrProProValThr.....ArgArgArg 1001
990 CGCAGCAAGAGCTGTCGGCTGGG.TTGGCCCGCAGCGGACCAATAC 1038
1002 SerArgSerArgThrProProValThrArgArgSerArgThr 1018
1039 TCCA.....TCAACGGGTACGACCTCGCCATTTTCT 1070
1018 rSerProValThrArgArgSerArgThrSerProValThr 1035
1071 GAAAAACAACCTTCAAGTTCAAGACAGCCCTCAACGGTGGCGACCGG 1120
1035 rgArgArgSerArgThrSerProValThr.ArgArgArgSerArg 1051
1121 CCATGGTGGGATTTGCTTACGAGCGGTAAATCGCGCTAGACATCTCTG 1170
1051 gSerArgThrProProAlaIleArgArgSerArgThrProL 1068
1171 CCTACCTGCTTTTGGCGGATTTAATCGTGGGATACGACAGCGCGCA 1220
1068 euLeuPro.....ArgLysArgSerArgSerArgSer 1078
1221 AGCATGGGTGCTTGAATTGACGAGGAAGACCTCGC 1259
1079 ProLeuAlaIle.....ArgArgArgSerArg 1087
seq_name: sp_human:Q9UQ35

seq_documentation_block:
ID Q9UQ35 PRELIMINARY; PRT: 2752 AA.
AC Q9UQ35;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RNA BINDING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, complete cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016092; BAA83718.1; -;
DR InterPro; IPRO02965; P-rich_extensions.
DR PRINTS; PRO1217; PRICHEXTENSN.
SQ SEQUENCE 2752 AA; 299672 MW; 109C64F181097123 CRC64;

alignment_scores:
Quality: 116.50 Length: 447
Ratio: 0.613 Gaps: 19
Percent Similarity: 42.506 Percent Identity: 24.161
alignment_block:
US-09-303-518D-127 x Q9UQ35 ..
Align seg 1/1 to: Q9UQ35 from: 1 to: 2752

12 CAAAAAGGCTTAACCTCCCATCGCGGACAGCCGAGCAAGTCAATT 61
1717 ArgSerArgThrProProAlaIleArgArgSerProSerValSerPr 1733
62 ATGACGGGCGCGTCAATTACCGAAGTCGCGTTCGTTGGCGAAGATATGCC 111

749 ArgSerArgThrProProAlaIleArgArgSerProSerValSerPr 765
62 ATGACGGGCGCGTCAATTACCGAAGTCGCGTTCGTTGGCGAAGTATGCC 111
765 oGluProAlaGlu.....LysSerArgSerArgArgArgSerA 780
112 GGTATGCGCCCGTNGATGAAGTCAAGGAAGCGATGCGCTCAAAAAAGG 161
780 laSerSerProArgThrLysThrSerArgArgGlyArgSerProSer 796
162 CCAAGTCTCTTTGAAGACAAAAAGNATCCGGCGGTGCTTACCGCGC 211
797 ProLysPro.....ArgGlyLeuGluArgSer 805
212 CNGTTTCAGGCAAAATCGCGCCATCCATCGCGGCGGAAAAAGCGCTACTT 261
805 r..... 805
262 CAGTCGGTCTGTGATGCGTTGAAGCAAGCAAGCAATCGATTGCAAGC 311
806ArgSerArgSerArgArgGluLysThrArgThr 816
312 CTAGCGCCCGAAGCGTTGCAACACITTAACGCGCGGANGAANTNNGNCA 361
817 ThrArgArgArg.....ArgSerGlySerSerG 827
362 ATCTGATCCAATCGCGTTTGTGGACTGCGCTGCTANCCGTCGTTTCAGC 411
827 nSerThrSerArg.....Arg 834
412 AAATCCCTGCGTGCATGCGGACCGCTTCCGATCTTCGTCAATCGAT 461
834 InArgSerArgSerArgValThrArgArgArgGlyGlySer 850
462 GCACCAATCCGTNGC..... 479
851 GlyThrHisSerArgSerProAlaArgGlnGluSerSerArgThrSerSe 867
480GCCAGA.....CCTGTGTTGTGATCAAGAA 507
867 rArgArgArgArgGlyArgSerArgThrProProThrSerArgLys 884
508 GCGGCGGANGATTCAGACGANGTGTGGTATTTAGCCCTTTGACCGA 557
884 eArgSer.....ArgThrSerProAla.....ProTyrLysArg 895
558 GCGTAATCCATGCTGTAGGCGAGCTGGCGCAGAGTGCCTCTGAAA 607
896 Ser.....ArgSerArgAla.....Se 901
608 ATGCTGCCAATCGAAGACATGAATTCGGCGCGCCCGCATCCGCGCGGT 657
901 rProAlaThrHisArgArgSerArgThrProLeuLysSerArg. 917
658 TTGAGTGGCAGCAGACATTCATTTCAATTGAGCGGTGCGTCAACAAAC 707
918ArgArgSerArgSer 922
708 CGTTTGGACCATCAATATCAAGATGATTAATGCCATCGGAGCTTTGTTG 757
923 ArgThrSerProValSerArgArgSerArgSerArgThrSerValTh 939
758 CAACAGCGCTGTGAACACCGGCGGTGATTTGGTGGTCTCTCAA 807
939 rArgArgArgSerArgSerArgAlaSerProValSerArgArgSerA 956
808 GTCAACAAACACCGCTCTTTCGCTACCGTTTGGTGGCGAAGATATGCCA 857
956 rgSerArgThrProProValThr..... 963
858 AATTACTGCGGGCAATTGTTGACGACAGCAACCGCGTATTCGGTT 907

```

1733 oGluProAlaGlu.....LysSerArgSerArgArgArgSera 1748
112 GGTATGCGCCCTNGATGAAGTCAAGGAAGGCGATCCGTCACAAAGG 161
1748 laSerProArgThrLysThrSerArgArgGlyArgSerProSer 1764
162 CCAAGTGTCTGTTTGAAGACAAAAGNATCCGGCGTGGTGTACGGCC 211
1765 ProLysPro.....ArgGlyLeuGlnArgSe 1773
212 CNGTTTCAGCAAAATCGCGCCATCCATCGCGGCAAAAGCGCTACTT 261
1773 f..... 1773
262 CAGTGGTGTGATTGCGTGAAGGCAAGCAGCAAAATCGAGTTCGAACG 311
1774 .....ArgSerArgSerArgGlyLysThrArgThr 1784
312 CTACGCGCCGGAAGCCTTGGCAAACTTAAGCGCGCGANGAANTNNGCA 361
1785 ThrArgArgAsp.....ArgSerGlySerSerG1 1795
362 AUCTGATCAATCCGGTGTGTGACTGCGCTCGGTANCCGTCCTCAGC 411
1795 nSerThrSerArgArg.....ArgG 1802
412 AAAATCCCTGCGTGTGATCGGAGCGGTCGCAATCTCGCAATCGAT 461
1802 InArgSerArgSerArgSerArgValThrArgArgArgGlyGlySer 1818
462 GGACACCAATCCGPNCG..... 479
1819 GlyTyrHisSerArgSerProAlaArgClnGluSerSerArgThrSerSe 1835
480 .....GECAGA.....CCTGTGGTGTGTGATCAAGAA 507
1835 rArgArgArgArgGlyArgSerArgThrProThrSerArgLysArgS 1852
508 GCCGCGCAGGATTTTCAGACGANGTCTCGTATGAGCGGTGTGACGA 557
1852 erArgSer.....ArgThrSerProAla.....ProTPlsArg 1863
558 CCGTAATCAATGATGTGTAAAGCAGCTGCGCAGACGTGCGCTGAAA 607
1864 Ser.....ArgSerArgAla.....Se 1869
608 ATGCTGCCAATCGAACAACATGATTCGGCGGCCCGCATCCGCGCGT 657
1869 rProAlaThrHisArgArgSerArgSerArgThrProLeuIleSerArg 1885
658 TTGAGTGGCAGCACATTCATTTGAGCGCGTGGTGGTCAAAACAAAC 707
1886 ..... 1886
708 CGTTTGGACCAATATTCAGATGTAATGCCATCGGACGTTGTTTG 757
1891 ArgThrSerProValSerArgArgArgSerArgSerArgThrSerValTh 1907
758 CAACAGCGCGTGTGAACACCGGCGGTGATTCCTTGGTGGTGTCTCA 807
1907 rArgArgArgSerArgSerArgAlaSerProValSerArgArgSera 1924
808 GTCAACAACACCGCTCTGCGTACCGTTTGGTGGCAAGTATCGCA 857
1924 rgSerArgThrProProValThr..... 1931
858 AATTACTGCGGCGAATTGGTGTGACGACAGCAACCGGTGATTTCCGTT 907
1932 .....ArgArgArgSerArgSerArgThrPr 1940
908 CGGTATTGAAGCGCGAT.....TACACAAGCGCG 939
:|||||:

```

```

1940 oThrArgArgArgSerArgSerArgThrProProValThrArgArgA 1957
940 CACGATTATTGGACGCTACCAACATCAGATTCCTGTTATCGAAGAG 989
1957 rgSerArgSerArgThrProProValThr.....ArgArgArg 1969
990 CCGCAGCAAGAGCTGTTCGGTGGG.TTCGCGCGCAGCGGCAAAATAC 1038
1970 SerArgSerArgThrProProValThrArgArgSerArgSerArgTh 1986
1039 TCCA.....TCACGCTAGCAGCTCGGCCATTCTCT 1070
1986 rSerProValThrArgArgSerArgSerArgThrSerProValThra 2003
1071 GAAAAACAACCTTCAAGTTCACGACGCGTCAACGCTGGCGACCGC 1120
2003 rgArgArgSerArgSerArgThrSerProValThr.ArgArgArgSerAr 2019
1121 CCATGTGCGCGATTGCTACTTACGAGCGGTAAATCCCGCTAGACATCTG 1170
2019 gSerArgThrProAlaIleArgArgSerArgSerArgThrProL 2036
1171 CTACCTCTCTTTTTCGGCGATTAAATCTCGCGGATACCGACACGCGCA 1220
2036 euLeuPro.....ArgLysArgSerArgSerArgSer 2046
1221 AGCATTTGGTGTCTTGAATTTGACGAGAGAGACCTCGC 1259
2047 ProLeuAlaIle.....ArgArgSerArg 2055
seq_name: sp_invertebrate:Q9U0V2

seq_documentation_block:
ID Q9U0V2 PRELIMINARY; PRT: 2207 AA.
AC Q9U0V2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POSSIBLE MUS308 HOMOLOG (FRAGMENT).
GN L8342.01.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Masuy D., Purnelle B., Goffeau A., Ivens A.C., Lawson D., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL122012; CAS58415.1;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR SMART; SM00490; HELIC_C; 1.
DR ATP-binding; Helicase.
KW NON_TER 2207 2207
SQ SEQUENCE 2207 AA; 229262 MW; E461565DA0017B0C CRC64;

```

alignment_scores:

```

Quality: 116.00 Length: 447
Ratio: 0.611 Gaps: 21
Percent Similarity: 42.506 Percent Identity: 22.819

```

alignment_block:
US-09-303-518D-127/rev x Q9U0V2

Align seg 1/1 to: Q9U0V2 from: 1 to: 2207

```

1232 CAACCAATGCTTCGGCGCTGCGGTATTCGGCGAGCATTAATAATCGCGCAA 1183
|||||: ||| |||||: |||
813 GlnProArgLeuGlyGlnLeuSerValAlaThr... 824
1182 AAGCAGGGTAGGCAGGATGCTAGCGGATACGGCG...T 1145
:||||: |||: |||: |||: |||
825 .....ThrGlyGlnAlaSerAlaGlySerAlaArgAsnValAlaAla 839
1144 CGTAAGTACCAATCGCCACCATGGCGCGTCCGCCACCGTTGACGGCTGC 1095
|| :||: |||: |||: |||
839 erValAlaGluProAlaThrProValSerSerProProSerGluAla... 854
1094 GTGAACCTGAAGAGTTGTTTTCAGAAATGCCGAGCGTCTACGCGT 1045
854 ..... 854
1044 GATGGAGTATTTCGGCGTCCGGCGCAACACGCGAAGCATCTTTTCG 995
|||||: |||: |||: |||: |||
855 .....CysAlaThrThrHisPro...ProAspLeu 864
994 TCGCGCTTCTTCGATACCGAAATCTGATGTGGTAGCGTCCCAATAA 945
|||||
864 rLeuProSerLeuProThrGluHis..... 872
944 TCGTCCGGCGCTTGTGTAATCGCGCGTTCATACCGAACCGAAATCAC 895
|||||
873 ValSerProProCys.....AspProLeuLeuMe 882
894 GCGGTTGTCTCGTCAACCAATTCGCCCGCAGTAATTTGGCATACTTCG 845
|||||: |||: |||: |||: |||
882 targValSerSerGlyGly..... 889
844 CACCCAAACGGTCGACGAGAGGGTGTGTTGACTTGAGAACACCC 795
890 .....LysAlaGly..... 893
794 AAAGCAATCACGCGTCTGCTGCTAGCGCGCTGTTCAACAAACGCTC 745
:|: |||: |||: |||: |||
894 ...CysAlaAlaGlnSerValPheArgMetGlyValValAlaProThrPr 909
744 GATGGCAATTACATCTTGATAATTGATGTCCAAACGGTTTTGTTGCAC 695
| :||: |||
909 oThrSerLeuGlySer.....AspValLeuSerSera 920
694 CGACCGGCTCAATGAATGAATGTCGTCGCCACTCAACCGCGCGGATGC 645
|||||
920 laThrGly.....ValSerAlaLeuSerAlaAlaAsnAla 931
644 GGGCGCGCAATTCATGCTGTTTCGATGTGGCA..... 612
932 AlaProLeuSerAspLeuHisValThrAlaLeuProTyrLeuAlaTh 948
||||| ||| :||: |||
611 .GCATTTTTCAGCGGCGCGCTCTCGCGCAGCT..... 582
948 rAlaAlaAlaGlySerGlyAlaProAlaProAlaProGlyA 965
581 .....GCCTTACACATGGATTTTACGCTCGGTCAACCGGCTC 543
:|||||: |||: |||: |||: |||
965 rgThrCysPheThrLeuHisSer....AlaAlaArgThrThrGlyThrLeu 980
542 AATACGACACNACNCTGTTGAATCTCNCGGCTTCTTTGATCACAC 493
:|||||: |||: |||: |||: |||
981 SerSerSerAlaAla.....AlaThrSerThrGluGluProThrThrAl 995
492 CACAGGCTCTCGCGNACCGATTTGGTCTCATCGCATTCGACGAAGATGG 443
:|||||: |||: |||: |||: |||
995 aAlaAlaSerAlaSerThrGlyAlaValProProAlaLeuThrGluAla 1012

```

seq_name: sp_human:Q96HAL

seq_documentation_block:
ID Q96HAL PRELIMINARY; PRT; 984 AA.
AC Q96HAL;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:3792).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LYMPH, AND LYMPHOMA;
RA Strausberg R.; 2001) to the EMBL/GenBank/DBJ databases.
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008794; AA08794.1; .
SQ SEQUENCE 984 AA; 99036 MW; 36DEDB4419EEA8E6 CRC64;

alignment_scores:
Quality: 113.50 Length: 441
Ratio: 0.565 Gaps: 18
Percent similarity: 45.578 Percent Identity: 21.995

alignment_block:
US-09-303-518D-127/rev x Q96HAL

Align seg 1/1 to: Q96HAL from: 1 to: 984

1313 TTACGACACAGCGCGCTATTCGTTTCCCGGCGACGACGAAGCTGCA 126

```

alignment_scores:
    Quality: 112.00      Length: 468
    Ratio: 0.563        Gaps: 17
    Percent Similarity: 42.521    Percent Identity: 22.436

alignment_block:
US-09-303-518D-127/rev x Q99K60 ..
Align seg 1/1 to: Q99K60 from: 1 to: 631

1343 CAGCCCTTCCTCTCAANGTTTTCAGCACCTTA.....

```

```

111 ... |||.....:|||||
35 GlAspThrSerSerSerLeuAlaSerThrThrProValHisSerSe 51
1308 CAACAGCGGCGCCTATTCGTAATTCGCGCGCAGACAGAGCTGCACAAAG 1259
||||| ||| ... |||.....:|||||
51 rAsnSerAspProAlaThrArgProGlyAspSer..... 63
1258 CGAGGTCTCTTCGTCCTCAATTCACCAAGCAACCAATGCTGCGCGCTGTCG 1209
...:
64 .....:Thr 64
1208 GTATCGCCGACGATTAAATCGCGCAAAAGAGGAGGTAGCAGGATGCTAG 1159
|||||: |||.....:|||||
65 SerSerProValGlnSerSerThrSerProAlaThrArgAlaProG1 81
1158 CGCATTCAGCGCTCGTAAGTACCAATCGGCACCATGCGCGGTGCGCAC 1109
: |||.....: |||.....: |||.....: |||.....:
81 uAspSerThrSerThrAlaValLeuSerGlyThr.....SerSerProA 96
1108 CGTTGACGGCTGTCGTGAATTCGAAGAGTTGTTTTCAGGAATGGCGG 1059
||||| |||.....: |||.....: |||.....: |||.....:
96 laThrThrAlaProValAsnSerAlaSer..... 105
1058 AGGTCGTACGCGTGTAGTATGTCGCGCTGCGCGCAACCCAGCC 1009
...: |||.....: |||.....: |||.....: |||.....:
106 .....:SerProValAlaHisGlyAspThrSerSerPr 116
1008 GAACAGCTCTTCTGCGCGCCTCTTCGATAACGGAAATCTGATGTGGT 959
| |||.....: |||.....: |||.....: |||.....:
116 oAlaThrSerLeuSerLysAspSer..... 124
958 AGGTCCTCAATATCGTGGCGCCTGTAATCGCGCGTCAATACC 909
...: |||.....: |||.....: |||.....: |||.....:
125 .....:AsnSerSerProValValHisSerGlyThrSerSer 136
908 GAACCGGAATACAGCGGTGTCTGCGTCAACCAATTCGCCCGCATAT 859
||| |||.....: |||.....: |||.....: |||.....:
137 AlaProAlaThrAlaProValAspSerThrSerSerProValValHi 153
858 TTGCGATCTTTCGCCACCAACAGGTACGACAGAGCGGTGTTGTTGA 809
|||||: |||.....: |||.....: |||.....: |||.....:
153 sGlyGlyThrLeuSerProAlaThr.....SerProGlyAspSer 168
808 CTTGAGAACCAACCAAGCAATACAGCGCTGCTGTTTCAAGCGCGCTGT 759
||| |||.....: |||.....: |||.....: |||.....:
168 hrSerSerProAspHisSerThrSerSer.....Pro 179
758 GCAACAAACGTCGCGATGGCAATACATCTGTATTAATGATGTCACAAAC 709
|||||: |||.....: |||.....: |||.....: |||.....:
180 AlaThrArgAlaProGluAspSerThrSer.....Th 190
708 GGTGTTGTTTCACCGACCGCTCAATGAATGAATGCGTGCCACTCA 659
|: |||.....: |||.....: |||.....: |||.....:
190 rAlaValLeuSerGlyThrSerSer..... 198
658 AACCGCGCGGATCGCGCGCGCGCAATTCATGTTTCGATGTTGGCA... 612
|||||: |||.....: |||.....: |||.....: |||.....:
199 ..ProAlaThrAlaProValAspSerThrSerSerProValAlaHis 214
612 ..... 612
215 AspAspThrSerSerProAlaThrSerProSerGluAspSerAlaSerSe 231
611 ....GCATTTTCAGACGCGCTGCGCGCGCTGCTACACATGGA 566
231 rProValValHisGlyThrSerSerProAlaThr..... 243
565 TTTTACGCTCGCTCAACGGCTCAATACCAACANACNCTGCTCAATCN 516
||| |||.....: |||.....: |||.....: |||.....:
244 .....SerProLeuArgAspSerThrSerSerProValHisSerSer 257
515 TCNCGCGCTCTTTCATCACACACAGGCTGTCGCCGNCAGC..... 474
: ||| |||.....: |||.....: |||.....: |||.....:

```

```

258 AlaSerIleGlnAsnIleLysThrThrSerAspLeuAlaSerThrProAs 274
473 ....GGATTGTTCCATCGCATTCGAGAGATGGCGACGCGCTCGCAT 428
...: |||.....: |||.....: |||.....: |||.....: |||.....:
274 pHisAsnGlyThrSerValThrThrThrSerAlaLeuGlySerAla 291
427 CGACGGCAGGATTTTCTGAACGCGACGNTACGACGCGCAGTCCACAAA 378
...:
291 hrSer..... 292
377 CGGATTCGATCAGATTCGNNCNCNNTTCGCGCTTAAGTTTGCAA 328
|||||: |||.....: |||.....: |||.....: |||.....:
293 ProAsp.....HisSerGlyThrSerThrThrThrAsnSerSe 305
327 CGCTTCGGCGGTAGCTTCGAACCTCGATTTCGTCGTCGCT...TCAA 281
...: |||.....: |||.....: |||.....: |||.....: |||.....:
305 rGluSerAspLeuAlaThrThrProValThrSerSerMetProPheSer 322
280 CGGCAATCAGCAGCAGTGAAGTAGCGCTTTTCGCCGCGATGGATGGCG 231
|||: |||.....: |||.....: |||.....: |||.....: |||.....:
322 hrThrLysValThrSerGlySerAlaIleIleProAspHisAsnGlySer 338
230 GCGATTTCGCT..... 219
339 SerValLeuProThrSerSerValLeuGlySerAlaThrSerLeuValTy 355
218 .GAACNGCGCGGTAAACACACCGCGGATNCCTTTTGTCTTCAACA 170
...: |||.....: |||.....: |||.....: |||.....: |||.....:
355 rAsnThrSerAlaIleAlaThrPro.....ValSerAsnG 368
169 GCATTGCGCTTTTTCGCGGATCGCTTCTGACTTTCATCNCAGGG 120
...: |||.....: |||.....: |||.....: |||.....: |||.....:
368 lyThrGlnProSerValProSerGlnTyProValSerProThrMeta 384
119 CGCATACCGCATATTTTCGCCAAGCAACGCGACTTCGTAATGACGGG 70
...: |||.....: |||.....: |||.....: |||.....: |||.....:
385 ThrThrSerSerHisSerThrIleAlaSerSerSerTyThrSerThVa 401
69 CCGG 66
401 lPro 402
seq_name: sp_plant:09SUVO
seq_documentation_block:
ID Q9SUVO PRELIMINARY; PRT; 857 AA.
AC Q9SUVO;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 96.9 KDA PROTEIN.
GN F8B4.120 OR A14G32420.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villaroel R., Gielen J., Van Montagu M., Hoheisel J., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Terry N., Ardiles W., Buysshaert C., Dasseville R., De Clerck R.,
RA De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villaroel R.,
RA Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;

```


RT	mucin."				
RL	Biochem. J.	308:89-96	(1995).		
DR	EMBL;	U12768;	AAC4852.1;		
FT	NON_TER			1	
FT	NON_TER			317	317
SQ	SEQUENCE	317 AA;	30360 MW;		
				491F505A8B8D0C93	07C6E.

alignment_scores:	
Quality:	109.00
Ratio:	0.599
Percent Similarity:	43.645
Length:	417
Gaps:	18
Percent Identity:	20.384

alignment_block:

US-09-303-518D-127/rev x Q29070

Align seg 1/1 to: Q29070 from: 1 to: 317

1283 CCCGGCAGACGAAGCTGCACAAGCGAGGTCCTCT..... 1248
 ||| ::||:: || | |||
 51 ProProGLUThrSerSeruicCluambgctm

1247TCGTCCAAATTCACGCAACCCAAATGCTTGGCGCTTCGGTATGCG 1202
|||||:||||: |||: ||| |||||: |||
67 nProSerSerSerSerAlaProThrThrSerAlaThrSerValGlnp 84

1201 CGACGATTAAATCGCGCAAAGCAGGGTAGGCAGGATGTCTAGCGGCATT 1152
||::: |||||

1151 ACGCGCTCGTAAGTACCAATCGGCACCATGCGCGGTGCCACGGTTGAC 1102
||||| |||||
89 SerLeuDecemb-mb 02
..... SerSerfgly... 88

1101 GGCTGTCGTGAACCTGAAGAGTTTGTTTTTCAGGAATGCCGAGGGTCG 1052
| : : : : : : : : : : : : : : : : |

```
93   iselAiaInrSerValGlnSer..... 100  
1051 TACGGCTGATGGAGTATTGTTCGGGTCCCGGCACCACCCCCC... 108
```

```

101 .....SerSerSerGlySerAlaProThrThrSerAla 111
      |||::: ::||| :::::

```

1001 TCTTTGCTCGGGCCTTCTTCGATAACGGAATCTGTATTGTGGTAGCGTCC 952
::: :::::|||||
112 ThrSerValGlnProSerSer

951 CAAATAATCGTGGCGGCCTTGTGTAATCGGCGCGTTCAATACCGAACCGG 902

IL9SerSerProPheSerSerThrIleSerValGlnPro. 132

901 AATTCACCCCGGGTCTGTCTGCCGCAGCAGCAAGGTTTGCTA

```
133 .....SerSerSerAlaproThr..... 140
```

851 ACTTTCGCACCCAAAACGGTAGCAAGAGCGGTGTTGACTTGAGA 802
||| ||| :|::: |:
141 ThrSerAlaThrGlyValGluLeuPheProGlnGluGluGluGlu

801 ACCACCAAGCAATCACGGCTCGGTGTTACAGCGCCTGTTGCAACA 752

152 152

152 152

701 TTTCACCGACCGGCTCAATGAATGAATGTGCGTGCCACTCAAACCGGC 652

651 CGGATCGGGCCGCCGAATTCATGTGTTTCGATGTTGGCAGCATTTTCAG 602

```

165 rSerSerSerProPLeSerThrIleSerValGlnProSerS 182
      .... ... ||| :|::: :: |||:

```

63 A 63
311 r 311

name: sp_bacteria:p94909

documentation_block:
PP94909 PRELIMINARY;
PP94000. PRT; 529 AA.

01-MAY-1997 (TRENBLrel. 03, Created)
01-MAY-1997 (TRENBLrel. 03, Last sequence update)
01-NOV-1999 (TRENBLrel. 03, Last sequence update)

HYPOTHETICAL 58.7 KDA PROTEIN (FRAGMENT).
Microbacterium ammoniaphilum.

Actinomycetales; Micrococcineae; Microbacteriaceae; Microbacterium; NCBI_TaxID=42460;

SEQUENCE FROM N.A.
STRAIN-ATCC 15354;

Striebel H.M., Seeber S., Jarsch M., Kessler C.;
Cloning and characterization of the MamI restriction-modification

Gene 172:41-46(1996).

hypocretin protein. 1 1

Mon Jul 1 09:25:31 2002

SQ SEQUENCE 529 AA; 58716 MW; 5DC1216237DF72D CRC64;

alignment_scores:
 Quality: 108.00 Length: 473
 Ratio: 0.548 Gaps: 24
 Percent Similarity: 41.649 Percent Identity: 24.101

alignment_block:
 US-09-303-518D-127 x P94909

Align seg 1/1 to: P94909 from: 1 to: 529

27 CCTGCCATCGCGGCGGACGAGCAAGTATATGACGGGCGCGTCA 76
 |||:|||||
 128 ProAlaArgArgGlyValLeuArgArgHisArgHisArgAlaArgG1 144
 |||:|||||
 77 TTACCGAAGTCGCTGCTTGGCGGAAGATATGCGGTATGCGCCCTNG 126
 |||:|||||
 144 yGluArgGlyArgGlyProArgGlnGlnValProArgGlnHisPro... 159
 |||:|||||
 127 ATGAAGATCAAGGAGGCGGATCGCTCAAAAGGCGCAAGTCTGTTGA 176
 |||:|||||
 160ArgGlyArgArgAspArgAlaGlyArgProGlyLeuHisArg 173
 |||:|||||
 177 AGACAAAAAGNATCC.....GGCGTGTGTTTACCGCGCNG 214
 |||:|||||
 174 ArgArgArgAlaArgArgArgArgGlnGlyArgProGlnValArgHisG1 190
 |||:|||||
 215 TTTCAGGCAAAATCGCGCATCCATCGCGGCGGAAAGCGCGTACTTCA 264
 |||:|||||
 190 yAspArgGlnHisArgAlaAspProArgArgProArgAspProArgAlaG 207
 |||:|||||
 265 TCGGTCTGTATTCGCTTGAAGCAACGACGAAATCGAGTTCGAACGCTA 314
 |||:|||||
 207 yHisHisProLeuArgGluGlyGlyGlnGlyGlnGluAlaArgProLeu 223
 |||:|||||
 315 CGCGCCGAGC.....GTTGGCAACCTTAAGCGCGGANG 349
 |||:|||||
 224 ArgAlaArgGlyGlyGlnGlyGlyProGlyGlyProAlaGlyArgHisPr 240
 |||:|||||
 350 AANTNNGNCAATCTGATCCATCCGTTTGTGGACTCGCGTGTGCTANC 399
 |||:|||||
 240 AlaLeuGlyAspGluAsp.....ArgP 248
 |||:|||||
 400 CGTCCGTTACGAAATCCCTCGCGTGCATCGCGAGCCGTTGCCATCTT 449
 |||:|||||
 248 roArgAlaArgGlnGlyAlaArgAlaAlaHisArgGluAlaLeuHisLeu 264
 |||:|||||
 450 CGTCAATCGCATGACACCAATCCGCTNCGCGCAGACCCCTGTGGTTGT 497
 |||:|||||
 265 ArgLeu.....GlnArgArgArgGlyArgAlaAspGlyCysAr 277
 |||:|||||
 498GATCAAGAAGCGGANGANGATTTCAGACGANGTTCGCTGTTG 543
 |||:|||||
 277 gAlaGlnGlyArgAlaArgArgAlaArgArgAlaGlyGlyGlyLeu 293
 |||:|||||
 544 ACCGCTTTGACCGAGCGTAAATCCATGTTGTGAAGGACGCTGCGCGAGA 593
 |||:|||||
 294 ProArgArgGluAspArg..... 299
 |||:|||||
 594 CGTGCGCTGTGAAATGTCGCAACATCGAACAACATGAATTCGGCGGCC 643
 |||:|||||
 300ValArgAlaHisArg.....P 305
 |||:|||||
 644 CGCATCGCGCGGTTTGAAGTGCAGCGACATTCATTTCATGACCGGTC 693
 |||:|||||
 305 roArgProArgGlyArgGlyAlaAlaArgLeuAspArgAlaGlyG1 321
 |||:|||||
 694 GTGCAACAAACCGCTTTGGACCATCAATATCAAGATGTAATTCGCAT 743
 |||:|||||
 321 yValGlyAlaArgProAlaArgProArgArgLeuArgHisProArgProA 338

744 CGGACGTTTGTTCGAACAGCGCGTCTGAACACCGAGCGCGTATTGCTT 793
 |||:|||||
 338 laAspLeu.ProHisGlyGlyPro..... 345
 |||:|||||
 794 TGGTGGTTCCTCAAGTCAACAACACCGCCTCTTGGCTACCGTTTGGGT 843
 |||:|||||
 346GlnGlyValAlaArgLeuAspHisProGlnGlyL 357
 |||:|||||
 844 CGGAAGTATCGCAATTACTGCGGCGGAATTGGTTGACGACGACACCG 893
 |||:|||||
 357 euGlu.....GlyAlaAlaGlySerArgArgHisPro 367
 |||:|||||
 894 CGTATTTCGCTGCGTATTGAA.....CGGCGCATTCACACAGCGC 937
 |||:|||||
 368 HisArgLeuArgAlaArgLeuHisGlnGlyArgGlyAspLeuLeuArg 384
 |||:|||||
 938 CGCAGATTTTGGGACGCTACCAATCAGATTCGCTTATCGAAGAA 987
 |||:|||||
 384 g.....ProAArgAspArgLeuGlyArgArg 394
 |||:|||||
 988 GCGCGCAGCAAGAGCTGTCGCGTGGGTTGCGCGCAGCGGACAAATA 1037
 |||:|||||
 394 lyProArgGlnGlyGln.....GlyAlaHisGlyGlyGlnGly 406
 |||:|||||
 1038 CTCATCAGCGTACGACCCCTCGGCCATTTCTGAAAAACAACTCTTCA 1087
 |||:|||||
 407 LeuArgHisAla..... 410
 |||:|||||
 1088 AGTTCACGACACCGCTCAACGGTGGCGACCG.....C 1119
 |||:|||||
 411GlyArgArgArgGlyValProSerGlnLeuAlaValLeuA 425
 |||:|||||
 1120 GCATGGT.....GCCGATTGGTACTTACGA 1145
 |||:|||||
 425 rglGlnGlyValCysGlnAlaSerAlaThrLeuAlaLeuTrpMetThrSer 441
 |||:|||||
 1146 GCCTGAATGCGCGTAGACATCCCTACCTGCTTTTGGCGGATTAA 1195
 |||:|||||
 442 GlyArg.....LeuLeuArgGluLeuAlaGlyLeuLeuCysArgAspSerV 457
 |||:|||||
 1196 TCCTCGCGCATACCGAC.....AGC 1215
 |||:|||||
 457 alGluSerHisLeuAspValThrTrpArgAlaAspAlaSerIleSer 473
 |||:|||||
 1216 GCGCAGCATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1265
 |||:|||||
 474 .ArgArgThrTrpThrValMetAsn...ThrArgLeuArgGlnAlaGlnL 489
 |||:|||||
 1266 CAGCTTCGTCGCGCGCAATATACGAATANGCCCGCTGTTGCGTAAGG 1315
 |||:|||||
 489 ysSerLeuAlaAlaArgArgProThrThrSerSerIleSerArgSer 505
 |||:|||||
 1316 TGCTGGAAA 1324
 |||:|||||
 506 CysTrpGln 508
 |||:|||||

seq_name: sp_mammal:Q9N1P0

seq_documentation_block:
 ID Q9N1P0 PRELIMINARY; PRT: 818 AA.
 AC Q9N1P0;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE SUBMAXILLARY MUCIN (FRAGMENT).
 GN BSM1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCB1_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-BREED ANGUS;
 RX MEDLINE=20223253; PubMed=10759843;
 RA Jiang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;
 RT "The central domain of bovine submaxillary mucin consists of over 50
 tandem repeats of 329 amino acids: chromosomal localization of the
 RL BSM1 gene and relations to ovine and porcine counterparts.";
 DR Eur. J. Biochem. 267:2208-2217(2000).
 EMBL; AF178428; AAF67279.1;
 FT NON_TER 1
 FT NON_TER 818 818
 SQ SEQUENCE 818 AA; 73014 MW; B5BB44F84F66F86B CRC64;

alignment_scores:
 Quality: 108.00 Length: 431
 Ratio: 0.535 Gaps: 25
 Percent Similarity: 46.868 Percent Identity: 24.594

alignment_block:

US-09-303-518d-127/rev x Q9N1P0 ..

Align seg 1/1 to: Q9N1P0 from: 1 to: 818

```

1301 GGGCCNTATTCGATTTCGCC...GGCAGCAAGAGCTGCACAAAGCGAG 1255
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
261 GlyProGlyThrSerAlaProSerGlyGluThrSer.....GluSerAr 275
1254 GTCTTCT.....TCGTCCAATTCACCAAGCAACCAAGCTTCGCG 1217
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
275 gSerSerValProGlyGlySerGluThrThrGlnGlnProGlyAlaGly. 291
1216 CGCTGTCGTTATCGCGACGATTAAATCGCGCAAAAGCAGGTTAGCAGG 1167
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
292 .....SerGluSerProThr.....GlySer 296
1166 ATGCTAGCGGCTATTCGCGCTCGTAAGTACCAATCGGACCAATCGCGG 1117
:|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||
297 LeuSerProGlyValThrArgThrThrAlaLeuArgGlySerGluThrAr 313
1116 GTGCCACCGTTGACGGCTGTCTGTAACCTTGAAAGATTGTTTTCAGCA 1067
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
313 gValProSer.....ThrGlyValSerGlyLeuPro..... 323
1066 AATGGCGGAGGTCGTACGCGTGATCGAGTATTGTCGCGGTGCGCGCA 1017
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
324 .....GlySer 325
1016 ACCAGCGCAACAGCTCTTTGCGCGCTCTCTGATCAACGAATCTG 967
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
326 ThrGlnGlyGlySerAlaAlaThrGlyGlySerGlyAla..... 338
966 ATGTGCGTACGTCGCAAAATATCGTGCCTGTTGTGTAATCGCGCGT 917
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
339 .....GlySerGlyPro.....ThrAlaProV 346
916 TCAATACCAACGGAATATCAGCGGTGCTGCTGCAACCAATTCGCC 867
:|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||
346 alSerGlyGluThrArg...ThrSerValIleSerGlyThrAsnValPro 361
866 GCAGTAATTGCGATACTTTCGACCCCAACAGGTACGCAAGCGCGTGG 817
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
362 Val.....SerGlyAlaProValThrProGlySerSerAlaG1 374
816 TTGTTGACTTGAGAACCAACCAAGCAATCACGCGCTCGGTGTTTCAGAC 767
| :||| :||| :||| :||| :||| :||| :||| :||| :|||
374 Y...SerSerGlyAlaProGlyThrGlyGlyProGlySerGluThrAla 390
766 GGCCTGTTGCAACCAACGTCGCGGCAATACATCTTGATAATGATG 717
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
390 exProLeuSerGlyAlaAlaGlyThrSerAlaThrGly..... 402
716 GTCCAAACGGTTTGTTCACCGACCGCGCTCAATGAAATGAATGCGGT 667
:|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||

```

```

403 ...SerGlyThrSerIleProProSerGlyAla..... 412
666 GCCACTCAAAACCG.....GCCGATCGCGGC 641
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
413 ProValThrProGluProProLeuIleSerThrGlyAlaSerAlaGlyP 429
640 CGCCGAATTCATGTGTTTCGATTTGCGAGCATTTTCAGACGCGCGTCT 591
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
429 roProAlaSerSerGluSerThrVal.....Thr 438
590 GCGCCAGCTGCCTTACACATGATGTTTACGCTCGGTCAACGCGCTCAA 541
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
439 LeuProGlyAlaThrGlyThrAspValLeuArgSerGlyThrSerLeuPr 455
540 TACCAGCANACNTCGTCTGAAATCNCNCGCGCTCTTTTGATCACAACA 491
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
455 oValSerGlyGlyAlaVal...ThrProAlaProSerProGlyGlySerS 471
490 CAGGCTCTGCCGCGNAGCGATGTTGTCGTCATTCGATTCAGGAAGATGCG 441
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
471 exAlaThrAlaGlyProGly...ValGlySerAlaThrThrValGlnAla 486
440 AACGGCTCGCATCGACGCGGATTTTGTGTAACGCGACGNTACGCGAG 391
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
487 SerGly...AlaThrGlyAlaAspValLeu.....ArgSe 497
390 CCAGTCCCAAAACCG...GATTGGATCAGATTGCNCCNANNNTTCGCG 344
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
497 rGlyThrSerLeuProValSerGlyValAlaValSerProGlySerSerP 514
343 CCTTAAAGTTTCCCAACGCTTCGCGCGGTAGCGTTTCAACTCGATTTCG 294
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
514 ro.....GlyArgSerGlyAlaThrAlaValSerSerGlnGly 526
293 TGTGTCCTTCAACGCGCAATCAGCACCGACTGAAGTACGCTTTCGCC 244
||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||
527 SerGlnProThrValAlaLeuSerGlyAlaThrGlyThrSerValGlyPr 543
243 G.....CGATGGATGCGCGGATTTTGCCTGAAACNGCGCGGTAA 203
| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||
543 oSerGlyThrArgPheSerSerSerAlaIleProAlaThrProGlySert 560
202 ACACC.....ACGCGCGGATNCTTTTCTCTTCA 174
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
560 hrThrGlyArgAlaAlaGlyAlaGlyThrProGlyValAspSerGln 576
173 AACGACCTTGGCCTTTTTCAGCGCATCGCTTCTTTCGACTTTCATCNA 124
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
577 ThrAlaSerLeuProAlaAlaAlaArgProThrAlaLeuGly..... 590
123 GGGCGCATACCGCATATCTTCGCCAAGCAACGCGACTTCG 81
||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||
591 .....ProGlyThrSerAlaProSerGlyGluThrSer 601

```

seq_name: sp_virus:Q39781

seq_documentation_block:

```

ID Q39781 PRELIMINARY; PRT; 866 AA.
AC Q39781;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MEMBRANE GLYCOPROTEIN.
OS Equine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10326;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HH1;
RA Kirisawa R., Kobayashi T., Kawakami Y., Iwai H.;
RT "Nucleotide sequences of open reading frames 1, 24 and 71 of an
attenuated equine herpesvirus-1.";

```

RP	SEQUENCE FROM N.A.
RX	MEDLINE=98198465; PubMed=9531546;
RA	Shah S., Tugendreich S., Forbes D.J.;
RT	"Major Binding Sites for the Nuclear Import Receptor Are the Internal
RT	Nucleoporin Nup153 and the Adjacent Nuclear Protein Tpr.";
RL	J. Cell Biol. 141:31-49(1998).
DR	EMBL; AF045567; AAC1273.1; -
DR	InterPro; IPR001876; Znf-RanBP.
DR	Pfam; PF00641; zf-RanBP; 5.
DR	SMART; SM00547; ZNF_RBZ; 5.
KW	Protin.
FT	
NON_TER	1
SEQ	1219 AA; 123248 MW; 24FRQFG68G3696P6 CDS64.

```

alignment_scores:
    Quality: 108.00      Length: 516
    Ratio: 0.448        Gaps: 28
    Percent Similarity: 46.705    Percent Identity: 21.124

alignment_block:
US-09-303-518D-127/rev x 073749 ..

Align seg 1/1 to: 073749 from: 1 to: 1219

1322 TCCGACACCTTACGACACAGCGGCCNTATTTCGTATTGCGCGGG..... 1278
|||||::: ::::: ::::: ||| |||
312 SerSerAsnThrLysSerThrValProLeuSerSerThrProGlyLeuGI 328

```

[illegible][illegible]

990 GCCTTCTTCGATACGGAATCTGATTGTGGTAGCGTCCCAATAATCGT 941
|||||:|
898 rProSerAla..... 901
940 GCGCGCTTGTCTAATCGCGCTTCAATACGACCGGAATACACGCG 891
:|:|:|:|
902 ..SerProGlnGluThrAlaAlaIleSerArgMetalGlnThrGlnArg 917
:|:|:|:|
890 TTGCTGCTGCTCAACCAATTCGCCCGCAGTA.....ATTG 856
:|:|:|:|
918 ThrGlyThrSerArgGlySerAspThrIleSerLeuAlaSerGlnAlaTh 934
:|:|:|:|
855 CGATACCTTTCCGA.....CCAAAACGGTACGCAAGAGCGGTGTTGT 812
|||||:|
934 rAspThrPheSerThrValProThr..... 943
:|:|:|:|
811 TGACITTGAGAACCCCAAGCAATACGCGCTCGTGTTCAGACGGCT 762
:|:|:|:|
944ProProSerIleThrSerSerGlyLeuThrSerPro 955
:|:|:|:|
761 GTTGCAACAAACGTCGCGATGGCAATTACATCTTGATAATTGATGTC 712
:|:|:|:|
956GlnThrGln 958
:|:|:|:|
711 AACGCTTTTTCACCGCAGCGCTCAATGAAATGAATGTCGTCGCCAC 662
|||||:|
958 nThrHisThrLeuSerProSerGlySerGlyLys..... 969
:|:|:|:|
661 TCAACACGCGCGGATCGGCGCCCGCAATTCATGTTTCGATGTTGGCA 612
:|:|:|:|
970ThrPheThrAla 974
:|:|:|:|
611 GCATTTTCAGACGCGAGCTGCGCGCAGCTGCCTTACACACATGGATTT 562
:|:|:|:|
975 LeuIleSerAsnAlaThrProLeuProVal..... 984
:|:|:|:|
561 ACCTCGCTCAACGCGCTCAATACACAGCANACNTGCTCTGAATCNCN 512
|||||:|
985ThrSerThrSerSerAlaSerThrGlyH 994
:|:|:|:|
511 CGGCTTCT...TTGATCAACACACAGGCTGTCGCGNACGCGATGGTG 465
:|:|:|:|
994 isAlaThrProLeuAlaValSerSerAlaThrSerAlaSerThrValSer 1010
:|:|:|:|
464 TCCATCCATTCAGAGATGGCGACGCGCTCGCATCGACGCGAGGAT 415
:|:|:|:|
1011 SerAspSerProLeuLysMetGluThrSerGlyMetThrThrProSerLe 1027
:|:|:|:|
414 TTTGCTCAACGCGGNTACGCGCAGCTCCACAAACCGGATGGATCA 365
:|:|:|:|
1027 uLysThrAspGlyGlyArgArgThrAlaThrSerProPro..... 1041
:|:|:|:|
364 GATTGCNNCANNANTCNCGCCGCTTAAGTTGCCAACGCTTCGGGCGG 315
:|:|:|:|
1042ThrThrSerGlnThrIle..... 1047
:|:|:|:|
314 TAGCGTTTCAACTCGATTCGCTGCTTCAACGCGCAATACGACCGCA 265
:|:|:|:|
1048IleSerThrIleProSerThrAlaMetHisThr... 1058
:|:|:|:|
264 CTGAAGTACGCGCTTTTCGCGCGGATGGATGGCGG.....ATT 224
:|:|:|:|
1059ArgSerThrAlaAlaProIleProIleL 1068
:|:|:|:|
223 TGCCTGAACACNGGC.....GCGGTAAC 201
:|:|:|:|
1068 euProGluArgGlyGluAlaIleGlnIlePheProSerSerProMetGly 1084
:|:|:|:|
200 ACCAGCGCGGATCTTTTCTCTTCAACACGACACTGGCGCTTTTGTGAC 151
:|:|:|:|
1085 GlnThrProGlyThrThrProSerSerSerGlyGlyProThrProPr 1101
:|:|:|:|
150 GGCATCGCTTCTTCTGACTTTTCATCNAGGGCGGCATACCGGCATATCTT 101

seq_name: sp_human:Q9GZV6
748 PheGlyLeuLysLysAspGlyGluSerLysProPheLeu 761
|||||:|
seq_documentation_block:
ID Q9GZV6 PRELIMINARY; PRT; 1187 AA.
AC Q9GZV6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MUC4 PROTEIN SPLICED VARIANT SV14 (MUC4 PROTEIN VARIANT VII).
GN MUC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREATIC TUMOUR;
RA Choudhury A., Moniaux N., Hollingsworth M.A., Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin splice variants in pancreatic adenocarcinoma";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=20381033; PubMed=10920259;
RX Choudhury A., Moniaux N., Winpenny J.P., Hollingsworth M.A.,
RA Aubert J.P., Batra S.K.;
RT "Human MUC4 mucin cDNA and its variants in pancreatic carcinoma";
RL J. Biochem. 128:233-243(2000).
DR EMBL; AJ400851; CAC14136.1; -;
DR EMBL; AJ400633; CAC10062.1; -;
SQ SEQUENCE 1187 AA; 121925 MW; 62E4C93F86443B8F CRC64;

alignment_scores:
Quality: 107.50 Length: 446
Ratio: 0.578 Gaps: 17
Percent Similarity: 41.704 Percent Identity: 21.076

alignment_block:
US-09-303-518d-127/rev x Q9GZV6 ..
Align seg 1/1 to: Q9GZV6 from: 1 to: 1187
1340 CCTTCTTCTCAANGTTTCCAGACCTTACGCAACAGCGGCCNTATTC 1291
|||||:|
810 ProSerSerSerGlyAlaSerGlyThrThrProSerGlySerGluGlyL 826
:|:|:|:|
1290 GTATTTGCCCGGCGACGAGGATGCACAAACGCGAGGTTCTTCTGTC 1241
:|:|:|:|
826 eSerThrSerGlyGluThr.....ThrArgPheSerSerAsp 839
:|:|:|:|
1240 ATTCCAAGCAACCAATGTTGGCGCTGTCGATCGCGCGAGATTAA 1191
:|:|:|:|
839 roSerArgAspSerHisThrThrGlnSerThrThrGluLeuSerAla 855
:|:|:|:|
1190 TCCGCAAAAGCAGGCTAGGACGATGCTAGCGGCTTACGCGGTCTGTA 1141
:|:|:|:|
856 SerAlaSerHisGlyAlaIleProValSerThrGlyMetAlaSerSer 872
:|:|:|:|
1140 AGTACCAATCGGCACCATGCGCGGTGGCCACCGTTGACGGTGTGCTGA 1091
:|:|:|:|
872 eValPro...GlyThrPhe.....HisProThrLeuSerGluAla... 884
:|:|:|:|
1090 ACTTGAGAGTTTGTTCAGGAATATGCCAGGCTGTCACCGCGATG 1041
:|:|:|:|
884 884
1040 GAGTATTGTTCGGCTGCGCGCAACCCAGCAGCTCTTTGCTGGG 991
:|:|:|:|
885SerThrAlaGlyArgProThrGlyGlnSerSerProThrSe 898

```

1101 oHisSerSerArgArgLeuAlaSerProLeuAlaProLeuSerVali 1118
100 CGCAACGACGCGGCTCGGTAAATACGGGCGCGTCA 63
1118 leProSerThrSerGlnThrMetAlaArgSerSer 1130
seq_name: sp_human:Q14395

```

```

seq_documentation_block:
ID Q14395 PRELIMINARY; PRT; 505 AA.
AC Q14395;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MUCIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=STOMACH;
RX MEDLINE=93194895; PubMed=7680650;
RA Toribara N.W., Robertson A.W., Ho S.B., Kuo W.L., Gum E., Hicks J.W.,
RA Gum J.R. Jr., Byrd J.C., Siddiki B., Kim Y.S.;
RT "Human gastric mucin. Identification of a unique species by expression
RT cloning.";
RL J. Biol. Chem. 268:5879-5885 (1993).
DR EMBL; L07518; AAB61945.1; .
FT NON_TER 1
FT NON_TER 505
SQ SEQUENCE 505 AA; 51967 MW; 45D7CB1E6050E07D CRG64;

```

```

alignment_scores:
Quality: 107.00 Length: 310
Ratio: 0.690 Gaps: 14
Percent Similarity: 50.000 Percent Identity: 22.258

```

alignment_block:

US-09-303-518d-127/rev x Q14395

Align seg 1/1 to: Q14395 from: 1 to: 505

```

926 ATTCGCGCGTCAATACGACCGGAAATACGCGGTGTCTCGGTCAAC 877
87 ValAlaProThrSerThr...ThrThrIleThrProAsnProThrSerTh 102
876 CAATTGCGCGCGCAGTAAATTCGATACTTTCGACCGCAACGCGTACGCA 827
102 rArgThrArgThrProValAlaHisThrAsnSerAlaThrSerSerArgp 119
826 AGAGCGGTGTTGTTGACTGTGAGAACCCCAAGCAATACGCGCGTCTG 777
119 roProProPheThrThrHisSerProThrGlySerSer..... 133
776 GTGTTACAGCGCGCTGTTCGAACACGTCGCGATGCGCAATTACATCTTG 727
134 .....PropheSerSerThrGlyProMetThrAlaThrSer.. 145
726 ATAATTGATGTCACAAACGGTTTGTTCGACCGCGGCTCAATCAAT 677
146 .....PhelysthrThrThrThrThrThrThrThrThrThrThr 157
676 GAATGTCGCGCCACTCAACCGCGCGGATGCGGCGCGGCAATTCATGT 627
158 .....ProGlnThrThrPro 162
626 GTTCGATGTTGCGAGCATTTTCAGACGCGACGCTCTGCGCGCGCTT 577
163 LeuThrHisValProProPheSerThrSerLeuValThrProIleThr.. 178

```

```

576 ACACACATGATTTTACGTCGCTCAACGCGCTCAATACCGACANA.... 531
179 .HisThrValIleThrProThrHisProGlnMetSerThrSerAlaTyri 195
530 .....CNTGCTCTGAAATCN 516
195 leHisSerThrProThrGlyThrIleAlaSerProThrThrVallysala 211
515 TCGNCGGCTTCTTTGATCACACACGAGGCTCGCGNACGCGGATTGGT 466
212 ThrArgSerThrTyThrAlaProLeuMetThrAlaThrThrSerArgil 228
465 GTCCATCGCATTCAGACGATGCGACGCGTTCGCGATCGACGCGAGGA 416
228 eSerGlnAlaHisSerSerIleSerThrAlaLysThrSerThrSer.... 243
415 TTTTGTGTAACGAGCGGNTACGCGGAGTTCGCGCTTAAAGTTGCAACGCTCGGGCGC 316
244 .....LeuHisSerHisAla...SerSerThrHisHisProGlu..... 255
365 AGATTGCNNCNCNANTTCNTCGCGCTTAAAGTTGCAACGCTCGGGCGC 316
256 .....ValThrProThrSerThrThrThrAsnValThrProLy 267
315 GTAGCGTTCG.....AACTCGA 299
267 sSerThrSerArgAspThrSerThrProValThrHisThrThrSerAlat 284
298 TTTTCGTCGTTGCTTCAACGCAATCAGACGCGACTGAAGT...ACGCGC 252
284 hrSerSerArgProProThrProIleThrThrHisSerSerProThrArg 300
251 TTTTCCGCGCGATGATGCGCGGATTTGCTCGTGAACGCGCGGTAAA 202
301 SerSerPro.....LeuSerSerThrGlyProMetTh 311
201 CACACGCGCGGATNCITTTTGTCTTCAACAGACACTTGGCCTTTTGA 152
311 rAlaThrSer.....IleLysThrThrThrThrTyPro.....T 323
151 CGGATCGCCTTCCTTCATCATCAGGCGCGCATACCGCATATCT 102
323 hrProSerHisProGlnThrThrLeuThrThrHisValProProPheSer 339
101 TCGCAACGACGCGACTTCGTTAATGACG 72
340 ThrSerSerValThrProSerThrHisThr 349

```

seq_name: sp_invertebrate:Q9V515

seq_documentation_block:

ID Q9V515 PRELIMINARY; PRT; 746 AA.

```

AC Q9V515;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG8181 PROTEIN.
DE CG8181.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceinliger S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,

```



```

:::||||:
406 SerSerSerSer.....SerSerSerPr 413
381 CAAACCGGATGGATCAGATGTCNNNANTTCNCGCGCTT..... 339
|||||:
413 oLysProPro.....AlaSerSerThrThrProLeuValThrL 426
338 ..AAGTTTGCACAGCTTGGCGCGCTAGCGTTTCGAACCTCGATTTCGTCG 291
|||:
426 yLysLysProThrThrThrAsnAlaProThrThrThrThrThrThr 442
290 TTGCTTCACGCAATCAGGACCGACTGAAGTACGCGCTTTTCGCGCGG 241
:::||||:
443 ThrThrThrThrThrThrThrThrProLysProThrArgArgThrLysPr 459
240 ATGGATGGCGCGGATTTCCTGAAACNGCGCGGTAAACACACGCGCG 191
|||:
459 o.....ProThrThrThrThrThrThrThrLysA 470
190 GATNCTTTTGTCTTCAACACAGACTTGGCCTTTTTCAGCGCATCGCT 141
:::||||:
470 laThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 486
140 TCCTTGACTTTCATCAGGCGCATACGCGCATATCTTCGCCAAGCAA 91
:::||||:
487 LeuLeuThr.....ThrGluProProThrSerAlaProLeuTh 499
90 CGGACTTCGGTAATGACGGGC 69
:::||||:
499 rThrThrThrThrLysLysThrGly 506

```

seq_name: sp_human:Q9Y2N2

seq_documentation_block:

```

ID Q9Y2N2 PRELIMINARY; PRT; 1294 AA.
AC Q9Y2N2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE WUGSC:H_D0953A04.3A PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL Genome Res. 8:1097-1108(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Cordes M., Bauer C., Holmes A.;
RT "The sequence of Homo sapiens PAC clone RP5-953A4.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006014; AAD28063.1;
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 1294 AA; 131942 MW; 6FC9761CC44162C2 CRC64;

```

alignment_scores:

```

Quality: 107.00 Length: 425
Ratio: 0.546 Gaps: 17
Percent Similarity: 46.118 Percent Identity: 22.824

```

alignment_block:

US-09-303-518d-127/rev x Q9Y2N2

Align seg 1/1 to: Q9Y2N2 from: 1 to: 1294

```

1115 TCGCCACCGTTGACGGCTGTGCTGAAC.....TTCAAGAG 1081
|||||:
561 SerProProThrSerLeuLeuAlaProSerThrAsnProLeuLeuGluSe 577
1080 TTTGTTTTTTCAGGAAATGGCGAGGTCGTACGCGTGATGAGTATTGT 1031
|||||:
577 rLeuLysLysMetGlnThrProProSerLeuProProCysProGluSerA 594
1030 CCGGCTGGCGGCAACCCAGCCGCAACAGC.....TCTTTG 996
:::||||:
594 laGlyAlaAlaThrThrGluAlaLeuSerProProLysThrProSerLeu 610
995 CTGCGGCTTCTTCGATAACGGAATCTGATTGTGTAGCGTCCCAATA 946
|||:
611 LeuProProLeuGlyLeuSerGln..... 618
945 ATCGTGGCGGCTTGTGTAAATCGCGCG.....TTCAATACCGAAC 905
:::||||:
619 ....SerGlyProProGlyLeuLeuProSerProSerPheAspSerLysP 634
904 CGGAATACACCGGCTTGTCT.....CGTCAACCAATTCGCCCGCA 864
|||:
634 roProThrThrLeuLeuGlyLeuLeuProAlaProSerMetValProAla 650
863 GTATTTTCGATACCTTCGACCCCAACAAAGTACGCAAGAGGGCGTTT 814
:::||||:
651 Thr.....AspThrLysAlaProProThrLeuGlnAlaGluThrAlaTh 665
813 GTTGACTTGAGAACCA.....CCCAAGCAATCACGCGCT 779
:::||||:
665 r.....LysProGlnAlaThrSerAlaProSerProAlaProLysG 679
778 CGGTGTTTCAGCGGCTGTTTGCACAAACGTCGCGATG..... 741
|||:
679 lnSerPheLeuPheGlyThrGlnAsnThrSerProSerProAlaAla 695
740 ...GCAATTACACTTCATAATTGATGGTCCAAACGGTTTGTTCGACC 694
|||:
696 ProAlaSerSerAlaSerProMetPheLysProIlePheThrAlaPr 712
693 GACCGGCTCAATGAAATGATGTCGCTGCCACTCAAAACCG..... 654
|||:
712 oProLysSerGluLysGluGlyProThrProGlyProSerValThra 729
654 ..... 654
729 laThrAlaProSerSerSerSerLeuProThrThrThrThrAla 745
653 .....GCCGATGGCGGCGCGCAATTCATGTGT 625
:::||||:
746 ProThrPheGlnProValPheSerSerMetGlyProProAlaSer...Va 761
624 TTCGATGTTGGAGCATTTTCAGACGGCAGCTGTGCGCCAGCTCGCTTAC 575
|||:
761 lProLeuProAlaProPhePheLysGlnThrThrThrProAlaThrAlap 778
574 ACACATGGATTTCACCTCGCTCAACCGCTCAATACAGCANACNTCOT 525
|||:
778 roThr..... 779
524 CTGAAATCTCNCGGCTTCTTTGATCACACACAGGCTCTGCCGNAG 475
:::||||:
780 .....ThrThrAlaProLeuPheThrGlyLeuAlaSerAlaThrSe 793
474 CGGATGTTGTCATCGCATTCAGCAAGATGCGCAACGCTCGGCTACG. 426
:::||||:
793 rAlaValAlaProIleThrSerAlaSerProSerThrAspSerAlaSerL 810
425 .....ACGGCAGGATTTTGTCTGACGGGACGNTAGCAGCCAGCTCCAC 381
:::||||:
810 ysProAlaPheGlyPheGlyIleAsnSerValSerSerSerVal... 825

```

```
380 AAACCGGATTCAGATCGATTCNNNNNNTTCGCGCGCTTAACTTGC 331
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
826 SerThrThrThrThrAlaThrAlaAlaSerGlnProPheLeuPheG1 842
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
842 yAlaProGlnAlaSerAlaAlaSerPheThrProAlaMetGlySerIleP 859
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
876 PheSerGlnSerLeuProThrAlaValProThrAlaThrSerSerSerAl 892
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
892 aAlaAspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAlat 909
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Smith H.O.,
251 TTTTCGCGCGATGATGCGCGCATTTTCGCTGAAACGCGCGGTAAA 202
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
876 PheSerGlnSerLeuProThrAlaValProThrAlaThrSerSerSerAl 892
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
201 CACCAAGCGCGATNTCTTTTGTCTCAACAGCAGCACT...TGGCCTTTT 155
DR EMBL; AE003500; AAF48533.1; -.
DR FLYbase; FBgn0030702; CG9056.
892 aAlaAspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAlat 909
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
104 TCTTCGCGCAAGCAGCGACTTCGGTAATGACGGCGCGTCATTAATGAC 55
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
926 AsnIleProPheGlySerSerAlaLysSerProLeuProSerTyrProG1 942
RA "The genome sequence of Drosophila melanogaster.";
54 TTGCTCCGCTCTGCCCGCGATGGC 30
RA Science 287:2185-2195(2000).
942 yAlaAsnProGlnProAlaPheGly 950
RA EMBL; AE003500; AAF48533.1; -.
seq_name: sp_invertebrate:09VXM5
seq_documentation_block:
ID 09VXM5 PRELIMINARY; PRT; 2406 AA.
AC 09VXM5;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE CG9056 PROTEIN.
GN CG9056.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
EX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin C., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster K.A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
```

```
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003500; AAF48533.1; -.
DR FLYbase; FBgn0030702; CG9056.
SQ SEQUENCE 2406 AA; 248803 MW; 740EAD78D684F76C CRC64;

alignment_scores:
  Quality: 107.00 Length: 505
  Ratio: 0.502 Gaps: 25
  Percent Similarity: 42.178 Percent Identity: 20.594

alignment_block:
US-09-303-518D-127/rev x 09VXM5 ..
Align seg 1/1 to: 09VXM5 from: 1 to: 2406
1298 CCNTAT.....TCGTATTTCGCGCGCAGCAGCAAGCT 1267
|||||
1266 ProTyrSerGlyLeuThrHisGlySerTyrLeuProProValLeuProVa 1282
1266 G.....CACAAAGCGAGCT 1253
1282 lAlaThrProAsnLeuSerAsnLeuProThrGlnHisArgSer...S 1298
1252 CTTCCTTCGTCCTCAATTCAGCAAA.....CCC 1227
|||||
1298 eAspSerArgAsnSerArgGluSerProAlaSerLeuLysSerThrPro 1314
1226 AATGCTTCGCGCTGCGGTATCG.....CCGACGATTAATCG..... 1188
|||
1315 SerAsnIleGlyLeuAsnValSerMetAlaProThrLeuArgSerIleTh 1331
1187 .CGCAAAGCAGGTAGGAGGATGTCTAGCGCATTACGCGCTCGTAAG 1139
|||||
1331 rProLeuAsnAsnSerSerAlaIleSerSerGlyAlaSerGlnProValV 1348
1138 TACCAATCGCACCATGCGCGCGTCCACCGTTGACGGCTGTCGTGAAC 1089
|||
1348 alSerValProSerAlaAsnSerThrAlaLeuSer..... 1360
1088 TTGAAGAGTTGTTTTCAGGAATAGCCCGAGGTGCTA.....CG 1048
1361 .....MetSerAsnProHisIleSerHisHisH 1371
1047 CGTCATGAGTATTGTCGCGC.....TGCGCGC 1019
1371 sValProAlaThrAlaSerGlyAlaPheSerSerSerAlaAlaGlyT 1388
1018 CAACCCAGCGCAAGAGCTTTTG.....CTCGCGCTTCTTCGATACG 975
|||
1388 hrSerThrProAsnSerGlyLeuSerThrLeuAlaValThrSerLeuSer 1404
974 GAATCTCATTTGTTAGCTGCCAAATAATCTGCGCGCTTCTGTGAAT 925
1405 Thr.....SerAlaAlaPro..... 1409
924 CGCGCGCTTCAATACCGAACCGGAAATCACGCGGTG..... 889
```

```

1410 .GlnProHisSerHisPheProGlnSerThrGlnMetLeuProGlnSerG 1426
887 .....
1426 lyAsnPheserSerValSerHisLeuLeuThrThrHisProMetSerSer 1442
878 ACCAATTGCGCGCCAGTAATTTGCGATCTTTTTCGCACCCCAAAACGGTACG 829
1443 GlnAsnGlnProMetValArgCysGlySer.....ThrLeuTy 1455
828 CAAGACGGGTGTTTGTGACHTGAGAACCCCAAGCAATCAGCGCT 779
1455 rSerGlnSerSerAlaAlaThrAlaProProSerAlaAlaAla.... 1470
778 CGGTGTTTCAGACGGCTGTGCAACAAACGTCGATGCAATTCATCT 729
1471 .....AlaValSerAsnPherProSerVal..... 1479
728 TGATAATTGATGTCACAAACGGTTTGTGTCACGACCGGCTCAATGAA 679
1480 .....LeuAlaValGlnSerLeuThrThrAlaValThrSerSer..... 1492
678 ATGAATGTCGTCGCCACTCAACCGCGGATCGCGGCGCGCAATTCAT 629
1492 .....
628 GTGTTTCGATGTTGCGCAGCATTTTCAGACGGCAGCTCTCGCGCAGCTGCC 579
1493 .....SerSerSerProSerThr 1498
578 TTACACACATGATTTTACGCTCGTCTC.....AAACG 547
1499 LeuSerSerSerValIleGlnLysValIleSerProLysGlyGluSerPr 4515
546 GCTCAATACCCACANACTCGTCTGAATCTGCGCGCTTCTTTGATCA 497
1515 oCysAsnLysAspArgAspSerSerTyrSerSerProAlaAsnAlaValV 1532
496 CAACCAACA.....GGTCTCGCC.... 480
1532 aThrThrCysAlaProThrThrProIleValSerSerGlySerAlaArg 1548
479 .....GCNACGGATTTGGTCTC 463
1549 ProThrProProLeuSerAsnCysThrSerMetGlyIleGlyMetValAs 1565
462 CATCGCATTCAGAGATGCGCAACGGCTCGCA..... 429
1565 nAlaAlaSerThrAlaArgSerSerCysAsnAlaIleSerProLeuSerI 1582
428 .....TCGACGGCAGGGATTTTGTGTAACGGACGGNTACGCGCAGTC 384
1582 leProAlaThrAlaGlyIleHisValSerAlaThr..... 1593
383 CACAACCGGATTCAGATTCAGATTCGNNCANNNTTCGCGCTTAAGTT 334
1594 ...AsnProSerPheGlnSerSerTyrPheProThrProLeuAlaPr 1609
333 TGCCAAACGCTTCGGCGCGTAGCTTCGAACCTCATTCGTCGTGCTT 284
1609 oProProSerSerProSerProAlaThrSerSerAlaAlaIleIleSerS 1626
283 CAACGGCATTCAGCACCGCTGAAAGTACGCGCTTTTCGCGC..... 243
1626 erSerAla.....SerGlnPheAsnProAlaValSer 1636
242 CGATGATTCGGCGGATTTTCCTGAACNGCGCGGTAAACACACCGCC 193
1637 HisSerMetSerSerIleValThrThrAlaGlyAlaThrThrThrAl 1653
192 CGGATNCNTTTTGTCTCAACACACCACTTGGCCTTTTTCAGGCGATCGC 143

```

```

1653 aserSer.....ValThrGlnP 1659
142 CTTCCTTGACTTTCATCAGGGCGCATACCGCATATTTCTTCGCCAAGC 93
1659 roSerValAlaAlaIleSerAsnProValThrAsnThrProHisProphe 1675
92 AACGGGACTTCGGTA 78
1676 SerAlaGluSerLeu 1680
seq_name: sp_bacteriap:Q9I3F4
seq_documentation_block:
ID Q9I3F4 PRELIMINARY; PRT: 352 AA.
AC Q9I3F4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HYPOTHETICAL PROTEIN PA1563.
GN PA1563.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OX Pseudomonas.
NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Brody L.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Smith K.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Reizer J., Saier M.H., Hancock R.E.W., Wu Z., Paulsen I.T.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL: AE004584; AAG04952.1;
SQ Hypothetical protein; Complete proteome.
SEQUENCE 352 AA; 40156 MW; 3030DSABF0513EBB CRC64;

alignment_scores:
Quality: 106.50 Length: 302
Ratio: 0.783 Gaps: 13
Percent Similarity: 45.033 Percent Identity: 22.848

alignment_block:
US-09-303-518D-127 x Q9I3F4 ..
Align seq 1/1 to: Q9I3F4 from: 1 to: 352
601 TCTGAAATGCTCCCAACATCGAAACACATGAATTCGGCGCGCCGATCC 650
21 AlaGluHisAlaAlaThrLeuGluIleProGlyTyrAlaLysSerLysPr 37
651 GCGCGGTTCAGTCGCGCAGCAGCATTCATTTCATT.....GACCGCGTCG 694
37 oAla.....SerAlaHisValGluPheValCysGlnAspAlaAspG 51
695 GTCAACAAACCGGTTTGGACCATCAATATCAAGATGTAATT..... 738
51 lyAlaGluArgLeuMetArgArgLeuArgPheAlaAspLeuIlePhePro 67
739 .....GC 740
68 ArgGlnTrpAlaArgGlyProGlyPheIleGluLeuProGluSerGlnAr 84
741 CATCGGACGCTTTGTTGCAACAGCGCGTCTGAACCGCGCGCGTATG 790
84 gIle.GluValLeuLeuAlaGluLeuAlaSerTyrProValCysGlySer 100
791 CTTTGGGTGTTCTCAAGTCAACAAACACCGCTCTTGGTACCGCTTTG 840

```



```

228 gHisSerGlyAlaHisLeuGlyGluLeuAspArgValAlaGlyVala 245
575 GTAAGGAGCGTGGCGAGAGCTCCGCTGCAAAATCTGCCACATCGAA 524
   : : : : : : : : : : : : : : : : : : : : : : : : : :
245 laLeuValAlaGlnAlaProValAlaGluGlu 255
625 ACACATGAATTCGGCGCCCGCATCCGCGGCTTTGAGTGGCAGC...CA 671
   : : : : : : : : : : : : : : : : : : : : : : : : : :
256 ..... 256
672 CATTTCAT...TTCATTCAGCGCGTGGTGCACAAACAAACCGTTGGACCA 718
   : : : : : : : : : : : : : : : : : : : : : : : : : :
263 svAlHisValValGluProAlaGlyArgGluAspAlaProAla 280
   : : : : : : : : : : : : : : : : : : : : : : : : : :
719 TCAATATCAAGATGTAATTCATCGGAGCGTTTGTTCACAGGCGCT 768
   : : : : : : : : : : : : : : : : : : : : : : : : : :
280 rglyHisGlyAspProAlaAlaValAlaLeu.....Glu 291
   : : : : : : : : : : : : : : : : : : : : : : : : : :
769 CTGAACACCGAGCGGTGATTCCTTGGGTGCTTCAAGTCAACAAACC 818
   : : : : : : : : : : : : : : : : : : : : : : : : : :
292 HisArgAlaAspHisProValAlaLeu...GlyAspGlnLeuAspGlnAr 307
   : : : : : : : : : : : : : : : : : : : : : : : : : :
819 AGCCCTCTTG.....CGTACCCTTTGGGTG 844
   : : : : : : : : : : : : : : : : : : : : : : : : : :
307 gGlyLeuAlaProAspArgAlaGlyAlaGlnArgAlaValGluGlnP 324
   : : : : : : : : : : : : : : : : : : : : : : : : : :
845 CGAAAGTATCGCAATTTACTCGGCGGAATGTTGACGACGACACCGC 894
   : : : : : : : : : : : : : : : : : : : : : : : : : :
324 roGlyArgGluArgProAlaGlyGlnValAlaAlaAspHisHis 340
   : : : : : : : : : : : : : : : : : : : : : : : : : :
895 GTGATTTCGGTTCGGTATTGAACGCGCGATTACACAAAGCGCGCAGA 944
   : : : : : : : : : : : : : : : : : : : : : : : : : :
341 .....AlaAlaAspAlaLeuSerGlyAlaAlaGlnAspAlaArgGlnAl 355
   : : : : : : : : : : : : : : : : : : : : : : : : : :
945 TTATTGGGACCGCTACCAACATTCATTC..... 974
   : : : : : : : : : : : : : : : : : : : : : : : : : :
355 aLeuAlaGlyLeuAlaArgAspGlnValHisProLeuValValArgThrG 372
   : : : : : : : : : : : : : : : : : : : : : : : : : :
975 .....CGTTATCGA.....AGAGCGCCGACGAAAGA 1001
   : : : : : : : : : : : : : : : : : : : : : : : : : :
372 lyAspArgHisArgAspArgCysLeuAspAlaArgProGlnGlnArg 388
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1002 GCTGTTCGGTGGTGGCGCGCAGCGGCAAAATCTCCATCAGCGTA 1051
   : : : : : : : : : : : : : : : : : : : : : : : : : :
389 Pro.....GlyLeuAlaGlnHisArgArgValGluArgLeuAlaLe 402
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1052 CGAC.....CTCGCGCATTTCTGGAACAAACAACTCTTCAAGTT..... 1091
   : : : : : : : : : : : : : : : : : : : : : : : : : :
402 uAspAlaAlaProArgGlyValAlaAlaArgGlnLeuArgValValG 419
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1092 .....CAGCAGCGCGTCAACGCGTGGCGACCGCGCAT 1124
   : : : : : : : : : : : : : : : : : : : : : : : : : :
419 lyValAlaAlaArgProHisGlnLeuGluArgArgProGlnHis 435
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1125 .....GTGTC.....GATTGGTACTTACAGCGCGTAA 1153
   : : : : : : : : : : : : : : : : : : : : : : : : : :
436 ProAspGlyLeuGlyAlaValProGlnGlnArgLeuProAlaGlyProSe 452
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1154 TCCCGCTAGACATCTGCTACCTGCTTTGCGCGATTTAATCGTCGCG 1203
   : : : : : : : : : : : : : : : : : : : : : : : : : :
452 rGlyArgArgHisProProPro.....ArgGlyArgSerArgA 466
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1204 GATACCGA.....CAGCGCGCAAGCATTTGGTTG...CTTGA 1238
   : : : : : : : : : : : : : : : : : : : : : : : : : :
466 rGlyAlaArgAlaCysArgAlaSerArgAlaSerValSerAlaGlyLeuAla 482
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1239 ATTGACGAAGAAGACCT 1256
   : : : : : : : : : : : : : : : : : : : : : : : : : :
483 AlaSerArgArgArgPro 488

```

seq_name: sp_virus:039782

seq_documentation_block:

```

ID 039782 PRELIMINARY; PRT; 867 AA.
AC 039782;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MEMBRANE GLYCOPROTEIN.
OS Equine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10326;
RN [1]
RP STRAIN=BK343;
RC STRAIN=FROM N.A.
RA Kirisawa K., Kobayashi T., Kawakami Y., Iwai H.;
RT "Nucleotide sequences of open reading frames 1, 24 and 71 of an
RT attenuated equine herpesvirus-1.";
RL J. Equine Sci 7:79-87(1996).
DR EMBL; D88734; BAA20038.1;
SQ SEQUENCE 867 AA; 86630 MW; C8DAD90D9B64407C CRC64;

alignment_scores:
    Quality: 106.50      Length: 460
    Ratio: 0.486        Gaps: 12
    Percent Similarity: 47.609      Percent Identity: 20.652

alignment_block:
US-09-303-518D-127/rev x 039782 ..

Align seg 1/1 to: 039782 from: 1 to: 867

1322 TCAGACACCTTACGCAACAGCGGCCNTATTCGTATTGCGCGGCAGAC 1273
   : : : : : : : : : : : : : : : : : : : : : : : : : :
50 SerSerSerProThrThrSerProProThrThrSerSerSerProThr 66

1272 GAAGCTGCACAAAGCGAGGTCTTCTGTCCTCAATTCACCAACCAATG 1223
   : : : : : : : : : : : : : : : : : : : : : : : : : :
66 rSerThrHisThrSerSerSerSerThrThrGlnSerSerSert 83

1222 CTGCGCGGTGTCGGTATCG...CCGACGATTAAATCGCGCAAGCAGG 1176
   : : : : : : : : : : : : : : : : : : : : : : : : : :
83 hrAlaAlaThrSerSerSerAlaProSerThrAlaSerSerThrThrSer 99

1175 GTAGGCAGGATGCTAGCGCATTCAGCGCTCGTAAGTACCAATCGGCAC 1126
   : : : : : : : : : : : : : : : : : : : : : : : : : :
100 IleProThrSerThrSerThrGluThrThrThrThrProThrAlaSe 116

1125 CATGCGCGGTGCGCCACCGTTCGCGTCTCGTGAACCTTGAAGAGTTGT 1076
   : : : : : : : : : : : : : : : : : : : : : : : : : :
116 rThrThrProThrThrThrAlaAlaProThrThrAlaAlaThrT 133

1075 TTTTCAGGAATGCGCGAGGTGCTACGCGTGATGAGTATTGTCGCGC 1026
   : : : : : : : : : : : : : : : : : : : : : : : : : :
133 hrThrAlaValThrThrAlaAlaSerThrSerAlaGluThrThrAla 149

1025 TCGCGCGCAACCCAGCGCGAGCTCTTCTCGCGCCTTCTCGATAAC 976
   : : : : : : : : : : : : : : : : : : : : : : : : : :
150 ThrAlaThrAlaThrSerThrProThrThrThrThrProThrSerThr 166

975 GGAATCTGATTGTGCTAGCTCCCAATAATCATCGTCGCGCCTTGTGTA 926
   : : : : : : : : : : : : : : : : : : : : : : : : : :
166 rThrThr.....ThrAlaThrThrThrValP 175

925 TCGCGCGTTCATACCGCAACCGGAATCAGCGGTGTCGCGTCAAC 876
   : : : : : : : : : : : : : : : : : : : : : : : : : :
175 roThrThrAlaSerThrThrThrThrThrThrThrAlaThrThrThr 191

875 AATTTCGCGCGCAGTAATTTCGATATCTTCGCACCCCAACG.....GT 832
   : : : : : : : : : : : : : : : : : : : : : : : : : :
192 AlaAlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrAlaAl 208

831 ACACAAGAGCGGTGTTGTTGACTTGAGAACCCCAACCAATTCAGC 782
   : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
208 aThrThrThraAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrT 225
781 GCTCGGTGTCAGACGGCTGTCGAAACAAACATCGGATGGCAATTACA 732
225 hrThraAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrThrT 241
731 TCTTGATTAATGATGTCACACGGTTTGTTCGACCGACGGCTCAAT 682
242 AlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrThrSerAl 258
681 GAAATGAATGTCGTCGACACCAACCGCGGATCGCGCGCGCGCAATT 632
258 aThrThrAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrT 275
631 CATGCTTTCGATGTCGACAGATTTTCAGACGGC..... 597
275 hrAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrThrT 291
596 .....ACGTCGCGCCAGCTGCCCTTACACACATCGATGATTTACGGCTC 556
292 AlaThrThrThraAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaTh 308
555 GGTCAACGGCTCAATACACAGCANACNCTGCTGAAATCNCNGCGGTT 506
308 rThrThrAlaAlaThrThrThraAlaAlaThrThrThraAlaAlaThrT 325
505 CTTTGATCACACACAGGCTGCGCGNAGCGGATGTCGTCATCGCA 456
325 hrAlaAlaThrThrThrGlySerProThrSerGlySerThrThrThrT 341
455 TTGACGACGATGCGAAGGCTGCGCATCGACGGCA...GGGATTTTCT 409
342 GlyAlaSerThrSerThrProSerAlaSerThrAlaThrSerAlaThrPr 358
408 GAACGACGCGNTACGACGCGAGTCCAAACCGGATTTGGATCAGATTGC 359
358 oThrSerThrSerThrSerAlaAla.....A 367
358 NNCNNATTCNTCGCGCTTAAGTTTGCAAGCTTCGGCGCGGAGCGT 309
367 laThrThrSerThrProThrProThrSerAlaAlaThrSerAla..... 381
308 TCGAATCGATTTGCTGCTGCTTCAACGGCAATCAGACGCGACTGAAG 259
382 ...GlySerThrThraAlaAlaProThrSerThrProThrThrAsp...Th 396
258 TAGCGCTTTTCCGCGCGATGATGCGCGGATTTGCTGAACNGCGG 209
396 rThrThrProSerGluAlaThrThrAlaAlaThrThrSerProGluSerThrT 413
208 CGGTAAC.....ACCACGCGC 192
413 hrValSerAlaSerThrThrSerAlaAlaThrThrThraAlaAlaThrThrGlu 429
191 GGATNCTTTTCTTCAACAGCACTTGGCTTTTTCAGCGCATCGC 142
430 SerHisThrSerProAspSerThrGlySerThrThrAlaGluPr 446
141 TTCCCTTGACTTTCATCAGGCGGCATACCGCATATTTCTTCGCAAGCA 92
446 oSerThrPheThrLeu.....ThrProSerThrAlaThrProSer. 460
91 ACGGCACTTCGGTAATGACGGCGCCCTCATATAATGACT..... 54
461 .....ThrAspGlnPheThrGlySerSerAlaSerThrGluSerAspSer 475
53 .....TGCTCCGGTCTGCCCGGATGGG 30
476 ThrAspSerSerThrValProThrThrGly 485

seq_name: sp_invertebrate:076602
seq_documentation_block:
```

1069 GGAATGGCCGAGGGTGGTACGGTGTGATGAGTATTG.....TCC-1029
468SerThrTyThrProSerSerSer 475
1028 GGCTGGGCGCAACCCAGCGG...AACAGCTCTTTGCTGGCGCTTCCTC 982
476 ThrAlaGlyAlaThrSerProAlaThrGlnSerThrLysProThrI 492
981 GATAACGGAATCTGATTGTGGTAGCGTCCCAAAATAATGTCGCGCGCTT 932
492 eGlyThrSerMet.....SerSerGlyProT 501
931 GTGTAATCGCGCGTTCATACCAACCGGAATACACGGGTGCTCGG 882
501 hrThrValAlaPro.....GlyAla 507
881 TCACCAATTCGCGCGCAGCTAATTTGCGACTATTTGCGACCAAAACGGT 832
508 SerThrGluSerThrValGlnSerSerThrProSerGlyThrThVa 524
831 ACGAAGAGCGGTGTTTGTGACTGAGACCCCAACCAACGATCAGCG 782
524 lThrLeuProSerGlySerSerThr..... 532
781 GCTCGGTGTCAGAGCGGCTGTGCAAAACACGTCGATGGCAATTACA 732
533AlaThrAlaGlyThrSerProGlnAlaSerThr 543
731 TCTTGATAATTGATGCTCCCAACGGTGTGTTGCGACCGCGCTCAAT 682
544ValThrThrValThrAspThrValSer..... 554
681 GAAATGAATGTCGTGCCACTCAACCGCGGATGCGGCGCGCGAATT 632
555GlySerT 557
631 CATGTGTTTCGATGTCGACGATTTTCAGACGCGACGTGCGCCAGCT 582
557 hrValThrSerGlnThrAlaGluSerSerLeuSerThrGluSerProThr 573
581 GCTTTACACATGATGATTTTACGTCGCTCAACCGCTCAATACCCAGCAN 532
574 SerAlaGlySer.....SerLeuSerThrValSerThrValSe 586
531 ACNCTCGT.....CTGAATCNCNCGGCTTCTTGTGATCACAA 494
586 rSerGlnProSerThrTyThrProValSerSerAlaSerSerIleTyrS 603
493 CCACAGGCTGTCGCGNAGCGGATTTGTCATCGCATTTGACGAGATG 444
603 erThrLeuSerGlySerThrGlySerThrAlaSerProGlyThrThrGlu 619
443 CGGAACGCTCGCATCGACGCGGAGGATTTGCTGAACGCGGNTACG 394
620 SerSerGlySerSerThrSerGlyProSerThrIleSerGlySerSerAl 636
393 CAGCGCATGCCAACACCGGATTCGATTCATGTCNCCNNTTCNCGC 344
636 aSerThrVal.....ThrGlySerThrV 644
343 CGCTTAAGTTTGCACCGCTTCGCGCGGAGGCTTCGAACTCGATTTCG 294
644 alThrGluAlaSerThrIleSerGlySerThrGluSerSerThrIlePro 660
293 TCGTTCCTTCAACGGCAATCAGACCGGCTGAAGTACGCGCTTTTCGCC 244
661 Gly.....SerThrGluSerThrValSerGluAlaSerThrValSerG 675
243 CGCATGATGCGCGGATTTGCTGAAACNCGCGCGGTAACACACCGC 194
675 ySerSerValSerThrValSerGlySerThr.....GluSerThrSera 690

193 CCGGATNCITTTTGTCTTCAACAGCACCTTGCCCTTTTGTGACGCATCG 144
690 laGlyAlaSerThrValSerGlySerThrGlySerThrValSerAspSer 706
143 CTTCTCCTGACTTTCATCAGGGGGGCATACCGGCATATTTCTTCGCCAAG 94
707 SerThrIleSerAspSerThrGly.....SerThrAsnAlaProG 720
93 CAACGCGACTTCGTAATGACGGCGCGTCATAAATGACTGCTCCCGT 45
720 ySerThrGluSerThrValThrGlySerSerValSerThrValSerGly 736
seq_name: sp_bacteriap:P73758
seq_documentation_block:
ID P73758 . PRELIMINARY; PRT; 574 AA.
AC P73758;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ABC TRANSPORTER.
GN SLR0864.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
Miyaoka N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Okumura S.,
Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996)
DR EMBL: D90909; BAA17810.1; -
DR InterPro: IPR003593; AAA.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran; 2.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_2.
KW Complete proteome.
SQ SEQUENCE 574 AA; 65027 MW; 2A75BC56BDA8FD17 CRC64;
alignment_scores:
Quality: 105.50 Length: 300
Ratio: 0.738 Gaps: 15
Percent Similarity: 47.667 Percent Identity: 23.333
alignment_block:
US-09-303-518D-127 x P73758 ..
Align seg 1/1 to: P73758 from: 1 to: 574
502 AAGAAGCCGCGANGAT...TTCAGACGANGTTCGTGTTGAGCGC 548
270 GlnGlnAlaPheValAspLysPheArgAlaSerAlaThrArgSerThrG 286
549 TTTTACCGAGCGTAAATCCATCGTGTAAAGCAGCTGGCGCAGAGTGC 598
286 nAlaLysSerArgLysGlnLeuGluLysIleGluArgValGluAlaP 303
599 CGTCTGAAATGCTGCCACATCGCAACACATGAATTCGCGCGCGCGCAT 648
303 ro.....ValGlyValArgThrLeuLysPheHisPhePro... 315
649 CCGCGCGGTTTGTAGTGGCAGCGACATTCATTCATTGACCGCGTGGTGC 698
316 ProAlaProArgSerGlyArgGlnValAlaLeuIleGluAspile..... 330


```
425 TCGA.....TGCCGAGCGGTTGCCATCTTC 450
|||||
654 gArgGlyArgSerArgThrProAlaArgGlyArgSerArgS 671
451 GTCATCGGATGGACACCAATCGCTNGCGGACACCCCTGTGGTGTGAT 500
|||||
671 erArgThrProAlaArgSerGlyArgSerArgSerArgThrProAla 687
501 CAAAGAGCGCGCGANGATTTCAGACGANGTNGCTGGTATTGACGCGTT 550
|||||
688 ArgArgGlyArgSerArgThrProArgArgGly..... 700
551 TGACCGAGCGCTAAATTCATGTGTGTAAGCGAGCTGCGCAGACGTCGCG 600
|||||
701 .....ArgSerArgSerArgSerLeuV 708
601 TCTGAAATGCTGCCACATCGAACAACATGTAATTCACATCGGACGT 750
|||||
708 alArgArgGlyArgSerHisSerArgThrProGlnArgArgGlyArgSer 724
651 GGCCGGTTTGAGTGGCAGCAGCATTCATTTCATTTGAGCGCGTGGTCAA 700
|||||
725 GlySerSer.....SerGluArgLysAs 732
701 ACAAAACCGTTGGACCATCAATTCACAGATGTAATTCACATCGGACGT 750
|||||
732 nLysSerArgThrSerGlnArgSerArgSerAsnSerSerProGluM 749
751 TTTTGTGCAACAGCGCTCTGAACACGAGCGGTGATTGCTTTGGTGG 800
|||||
749 etLysSerArgIleSerSerArgArgSer..... 761
801 TTCTCAAGTCAACACACCGCTCTTGGTACCGTTCCTGGTGGCGAAG 850
|||||
762 LeuSerSerProArgSerLysAlaLysSer..... 771
851 TATCGCAATTTACTCGGCGGATTTGGTGGACGACAGCAA..... 890
|||||
772 .....ArgLeuSerLeuArgArgSerLeuSerGlyS 782
891 .....CGCGTGATTTCCGGTTCGGTATTGAACGCGCGCATTC 929
|||||
782 erSerProCysProLysGln.....LysSerGlnThrPro 793
930 ACAAGCGCGCAGCATTTATTTGGGACGCTACCAATCAGATTTCGGTTA 979
|||||
794 ProArgSerArgSerGlySerGlnProLysAlaLysSerArgTh 810
980 TCGAAGAGCGCGCAGCAAGAGCTGTCGGTGGTGGCGCGCAGCGG 1029
|||||
810 rProProArg.ArgSer.ArgSerSerSerProProLysGlnLy 826
1030 GACAATACCTCAGCGTACGACCGCTCGGCCATTTCTGAAACAA 1079
|||||
826 sSerLysThrProSerArg.....GlnSerH 835
1080 ACTCTTCAGTTTCAGCAGCGCTCAACGGTGGGACCGCGCCATGTGC 1129
|||||
835 isSerSerSerProHisProLysValLysSerGlyThrProProArg 851
1130 CGATTGGTACTACGAGCGGTATGCGCTAGACATCTCCCTCCCTACCC... 1177
|||||
852 GlnGlySerIleThrSerProGlnAlaAsnGluGlnSerValThrProG 868
1178 .....TGCTTTTCGCGGATTTAATCTGCGCGATACCGACGCG 1217
|||||
868 nArgArgSerCysPhe.....GluSerSerProAspProGluLeuL 882
1218 GCAAGCATTTGGTGTGGAATTTGGACGAAGAAGACCTCGCTTTGTGCA 1267
|||||
882 ysSerArg.....ThrProSerArgHisSerCysSer 892
1268 GCTTCGTCT 1276
```

893 GlySerSer 895

seq_name: sp_bacteriap:Q99UM0

seq_documentation_block:

ID Q99UM0 PRELIMINARY; PRT; 435 AA.

AC Q99UM0;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE GLUCOSE-INHIBITED DIVISION PROTEIN GID.

GN GID OR S1094 OR SAV1251.

OS Staphylococcus aureus (strain N315), and

OS Staphylococcus aureus (strain Mu50).

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Staphylococcus.

OX NCBI_TaxID=158879, 158878;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);

RX MEDLINE=21311952; PubMed=11418146;

RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,

RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,

RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,

RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue K.-I., Kaito C.,

RA Sekinizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,

RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,

RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,

RT "Whole genome sequencing of methicillin-resistant Staphylococcus

RL Lancet. 357:1225-1240(2001).

CC -I- COFACTOR: FAD (BY SIMILARITY).

DR EMBL: AP003133; BAB42346.1; -

DR EMBL: AP003361; BAB57413.1; -

DR InterPro: IPR001327; FAD_pyr_redox.

DR InterPro: IPR002218; GIDA.

DR InterPro: IPR000205; NAD_binding.

DR InterPro: IPR001100; pyr_redox.

DR PRINTS: PR00368; FADPNR.

DR PRINTS: PR00411; PNDRTASEI.

DR ProDom: PD003738; GIDA; 1.

KW Complete proteome; FAD; Flavoprotein; Oxidoreductase.

SQ SEQUENCE 435 AA; 48371 MW; B76C254394051425 CRC64;

alignment_scores:

Quality: 105.00 Length: 272

Ratio: 0.784 Gaps: 13

Percent Similarity: 49.265 Percent Identity: 21.324

alignment_block:

US-09-303-518D-127 x Q99UM0 ..

Align seg 1/1 to: Q99UM0 from: 1 to: 435

331 GCAACCTTAAGCGCGGANGAANTNNGNCAATCTGATCAATCCGGTTT 380

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

10 AlGlyLeuAlaGlySerGluAlaAlaTyrGlnLeuAlaGluArgGlyI 26

381 GTGAGCTGCGCTG...CGTANCGTCCGTTTCAGCAAAATCCCTGCCGTCG 427

:|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

26 eLysValAsnLeuIleGluMetArgProValLysGlnThrProAlaHisH 43

428 ATCCGAGCGGTTTCGCATCTTCGTC.....AATCGATGGACACCAAT 471

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

43 isThrAspLysPheAlaGluLeuValCysSerAsnSerLeuArgGlyAsn 59

472 CCCTNGCGCGCACCCCTGTGTTGTGATCAAGAGCGCGCGANGATTT 521

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

60 AlaLeu...ThrAsnGlyValGlyValLeuLysGlu.....GluMe 72

522 CAGACGANGTNTGCTGTTGAGCGGTTTGACCGGCGTAAATCCATG 571


```

72 tArgLeuAsnSerIlelleGluAlaAlaAspLys.....85
572 TGTGAAGCCAGCTGGCGCAGACGTGCCGCTCAAAATGCTGCCAACATC 621
86 .....AlgaValProAlaGlyAlaLeuAlaVal 96
622 GAAACACATGAATTTCGGCGCGCCGCATCCGGCCGGTTTGAGTGGCACGA 671
97 AspArgHisAspPheSerGly.....103
672 CATTCATTTCATTCAGCGCGCTGGTGCACAACAAACCGTTTGACCATCA 721
104 ....TyrrIleThrGluThrLeuLysAsnHisGluAsnIleThrValIleA 119
722 ATTATCAAGATGTAATGCCATCCGA.....CGTTGTTTGCAACA 762
119 sn...GluGluIleAsnAlaIleProAspGlyThrIleIleAlaThr 134
763 GCCCTCTGACACACCGAG.....CGCGTATTCITTTGGGTGG 800
135 GlyProLeuThrThrGluThrLeuAlaGlnGluIleValAspIleThrGI 151
801 TTCTCAAGTCAACAACACCGCCTCTTCGCTACCGTTTGGGTGCGAAAG 850
151 LysAsp.....GlnLeuTyrrPheTyrrAspAlaA 161
851 TATCCCAAATAACTACTCGGGCGGAATTTGTTGACGAGACAAACCGCTGATT 900
161 laalAProIlelleGluLysGluSerIleAspMetAspLysValTyrrLeu 177
901 .....TCCGGTTTCGATTATTGAACGCGCGCATTC 929
178 LysSerArgTyrrAspLysGluAlaAlaTyrrLeuAsnCysPrometh 194
930 ACAAGCGCGCAGCATTTATTTGGACGCTACCAACATCAGATT.....972
194 rGlu.....AspGluPheAsnArgPheTyrrAspAlaValLeuGluA 208
972 .....972
208 lacGluValalaProValAsnSerPheGluLysGluLysTyrrPheGluGly 224
973 .....TCGGTTATCGAAGAGCGCGCAGCAAGAGCTGTT 1007
225 CysMetPropheGluValMetAlaGluArgGlyArgLysThrLeuLeuPh 241
1008 CGGCTGGTTGCGCG 1023
241 eGlyProMetLysPro 246
seq_name: sp_mammal:O18758
seq_documentation_block:
ID O18758 PRELIMINARY; PRT; 13288 AA.
AC O18758
AD 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SUBMAXILLARY APOMUCIN.
OS Sus scrofa domestica (domestic pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
[1] NCBI_TaxID=9825;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98070526; PubMed=9407109;
RT Eckhardt A.E., Timpte C.S., Deluca A.W., Hill R.L.;
RT "The complete cDNA sequence and structural polymorphism of the
RT polypeptide chain of porcine submaxillary mucin.";
RL J. Biol. Chem. 272:33204-33210(1997).
DR EMBL; AF005273; AAC62527.1; -.
DR InterPro; IPR000359; Cys_Knot.
```

```

770 AGACGGCTCTTTC .....AAACAAACGTCCTGCAATTAACATCTTG 727
    ||| |||
12546 ysThrGluAlaThrThrPheProGlyGluAsnGluThrThrArgValGly 12562
    ||| |||
726 ATAATTGATGGTCAACACGGTTTGTTCACCGACCGGCTC .....AA 683
    ||| ||| ||| ||| ||| |||
12563 IleAlaThrGlyThrThrGlyIleValSerArgLysThrLeuGluProG1 12579
    ||| ||| ||| ||| ||| |||
682 TGAATGAATGTCGGTGCACCTCAAAACC...GCCGGATCGCGCGCG 636
    ||| ||| ||| ||| ||| |||
12579 ySerTyrAsnThrGluAlaThrThrSerIleGlyArg.....12591
    ||| ||| ||| ||| ||| |||
635 AATTCATGCTTTTCGATGTGGACGATTTTCAGACGACGCTCTGCGCC 586
    ||| ||| ||| ||| ||| |||
12591 .....12591
585 AGCTGCTTACACATGATTTTACGTCGTCGAACG.....GC 545
    ||| ||| ||| ||| ||| |||
12592 SerGlyThrThrHisThrAspLeuProGlyGlyThrIleValLeuPr 12608
544 TCAATACACACANACNCTGCTGAAATCNCGCGCTCTTTGATCACA 495
    ||| ||| ||| ||| ||| |||
12608 oGlyPheSerHisSerGln.SerSerLysProGlySerSerValThr 12624
    ||| ||| ||| ||| ||| |||
494 ACCACAGGTCTGCCCGNAGCGGATGTGTCCATCGCA.....456
    ||| ||| ||| ||| ||| |||
12625 ThrProGlySerProGluSerGlySerGluThrGlyThrSerGlyGluPh. 12641
455 .TTGACAGAGTGGCGACGCGTCGCGCATCGACGCGAGGATTTGCTGA 407
    ||| ||| ||| ||| ||| |||
12641 eserThrThrValIleSerGlySerSerHisThrGluAlaThrThrPheI 12658
    ||| ||| ||| ||| ||| |||
406 ACGGAGGNTACGACGCGAGTCCCAACACCGGATTCGATTCAGATTGCNN 357
    ||| ||| ||| ||| ||| |||
12658 leGlyGlySerGlySerPro.....Gly 12665
356 CNNANTTCNCGCGCTTAAGTTTGCACGCTTCGGCGCGTAGCGTTC 307
    ||| ||| ||| ||| ||| |||
12666 ThrGlySerArgProGlyThrThrGlyGluLeuSerGly.....Th 12679
306 GAATCGATTTCTGCTGCTTCAACGCAATCAGACCGACTGAAGTA 257
    ||| ||| ||| ||| ||| |||
12679 rThrIleAlaSerGlyAsnAlaThrThrGluAlaThrThrSerThrGluT 12696
256 CGCGTTCGCGCGATGGATGGCGGATTTTCCTGAAACNGCGCG 207
    ||| ||| ||| ||| ||| |||
12696 hrArg.....IleGlyProGluThrGlyAla 12704
206 GTAAACACACGCGCGATNCTTTTGTCTTCAACACGACTTGGCCTTT 157
    ||| ||| ||| ||| ||| |||
12705 GlnThrThrValProGlySerGlnValSerGlySerGluThrGlyThrSe 12721
156 TTTGACGCGATCGCTCTTCTTACTTTCATCAGGCGGCGCATACCGCAT 107
    ||| ||| ||| ||| ||| |||
12721 rGluAlaValSerAsn.....ProAlaI 12729
106 ATTCTTCGCCAAGCAACCGACTCGGTGAATGACGGCGCGCTCA 63
    ||| ||| ||| ||| ||| |||
12729 leAlaSerGlySerSerThrThrThrThrSerGlyAlaSer 12743
seq_name: sp_invertebrate:Q9NK94
seq_documentation_block:
ID Q9NK94 PRELIMINARY; PRT; 581 AA.
AC Q9NK94;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 64.2 KDA PROTEIN.
GN BG:DS07295.4.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA Celnikier S., Rubin G.M.;
RA "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the Adh region.";
RL Genetics 153:179-219(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RA Celnikier S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nixhi K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA Zieran L.L., Rubin G.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003412; AAF44927.1; -.
KW FlyBase; FBgn0028861; BG:DS07295.4.
KW Hypothetical protein.
SQ SEQUENCE 581 AA; 64222 MW; 44841FDF16526701 CRC64;

```

alignment_scores:

Quality: 104.50 Length: 409
Ratio: 0.601 Gaps: 22
Percent Similarity: 42.543 Percent Identity: 23.472

alignment_block:

US-09-303-518D-127 x Q9NK94 ..

Align seg 1/1 to: Q9NK94 from: 1 to: 581

```

66 CGGCGCGCTCATACCAAGTCGCTGCTGGCGA.....AG 103
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
212 ArgLeuGluHisTyrArgAsnHisLeuArgTrpHisValThrProSerAr 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 AATATGCCGCTATCGGCCCTN.....125
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
228 gLeuGluAsnPheProProLeuGlyLeuAlaAsnThrAsnThrSerLysL 245
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 .....GATGAAGTCAGGAAGCGGATGCGCTCAAAAAGGCCAAGT 167
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245 euLeuValAspAlaAlaPheTrpLeuLeuLeuThrGlnLysGluLysTy 261
168 GCTGTTTGAAGACAAAAGNATCCGGCGTGGTGTTCACGCGCCNGTTT 217
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
262 LysLysLysAsnGlnIleArgSerGluLysAsnThrLeuLysIleGluAr 278
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 CAGGCAA.....AATCGCGGCATCCATCGCGCGGCGAAAGCGCGTACT 261
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 gGluGluIleLeuAsnLeuArgLysLysLeuSerLeuLysCysGlyGlu 295
262 CAGTCGCTGCTGATTCGCGTTGAAGCAA.....CGACCAATCCA 302
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 erCysGlyAsnAsnTrpArgIleArgLysGlyMetProSerArgAsnArg 311
303 GTTCGACCGCTACGCGCGCGGAGCGTGTGCAAACTTAAGCGCGGANGA 352
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 AlaValGlyLeuGlySerGlyGlyIleArgSer.....322
353 TNNNGNCAATCTGATCCCAATCGGTTTCTGGACTGCGCTGCGTANCCGT 402
||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
323 .....ValHisHisValAspProPheGlyAlaSerC 333
403 CCGTTCAGCAAAATCCGTCGCGTGCATGCCGAGCGGTCGCCATCTTCGT 452
333 ysProGlyGlnValAspCysGlyAsnCys..... 342
453 CAATCGGATGACACCAATCCGTCGCGGAGACCCCTGCTGTTGTATCA 502
343 .....AsnTrpLysThrAlaGlnGlyLys...ThrGlyGluAsnPh 355
503 AAGAAGCGCGGANGATTTCAGAGGANGTNTGCTGTTGATTGAGCGGTTG 552
355 eaAlaAlaArgMetGlyGlnThrGlyGlnAlaGlyCysAspGlnHisA 372
553 AAGAGCGCTAAATCCATGCTGTGAAGCAGCTGCGCAGAC..... 594
372 spLeu.Arg...IleHisLeuValValAlaAlaAspGluSerSerAspGI 387
595 .GTGCGCTGTAATGCTGCCAATCATCGAAACACATGAATTCGCGGCC 643
387 yClnProAsnGluAspAlaArgAspAspProAspHisGluAspGlyGly 404
644 CGCATCCGCGCGTTGATGTCAGCGACATTCATTTCATTGAGCGGTC 693
404 isGlyAlaAsnAsnLeuSer.....ProIle 412
694 GTGCA.....AACAAACCGTTTGGACCATCAATTA 725
413 ProAlaGluGlyHisProValLysAspLysSerGluTrpSerLeu..... 427
726 TCAAGATGTAATTCGATCGGACGTTTGTTCACACAGCGCTCTGAACA 775
428 .....IleAsnValGlyAlaValPhe.....TyrAsnV 437
776 CCGAGCGGTGATGCTGTTGGGTGTTCTCAA..... 807
437 alPhePheValPheTrpLeuGlyGlnSerGlnHisLeuAsnGlyArgGly 453
808 .....GT 809
454 GlySerAspGlnValGluGlyArgSerIleLysAlaGlnProArgProGI 470
810 CAACAAACACCGCTCTTCGTCACGTTTGGTGCAGAAAGTATCGCAA. 858
470 uAsnLysProSerHisIleArgThrSerGlnGlnGlnAspGlnGlnA 487
859 ..ATTACTCGGGCGAATTTGGTTGACGACAGCAACCGGCTGTTCCGGT 906
487 spGlnProAlaCysGlnLeuGlyAsnGlySer.....GlnSer 499
907 TCGGTATTGACGCGCATACAAAGCGCGACGATTATTTCGGGACG 956
500 AlaLysMetGlyGlyAlaValGlyGluGlyAla...ThrLysIleTrp... 514
957 CTACCAACATCAGATTCCGTTATCGAAGAGCGCGACAGAAAGAGCTGT 1006
515 .....ArgLysAlaAla..... 518
1007 TCGGTGGTTCGCGCGACCGGACAAATATCTCCATCAGCGGTACGACC 1056
519 .....AsnArgThrGluThrLysThrSerGlnArgThrArgPr 531
1057 CTGCGCCATTTCCTGAAACAAACACTCTTCAAGTTTCACGACAGCGCTCAA 1106
531 oLys.....GluArgGlyIleGlnArgGlnProGluT 542
1107 CGGTGGCGACCGCGCATGTC 1129
542 hr...GluThrAlaAlaThrCys 548
seq_name: sp_human:043419
seq_documentation_block:
```


